(54) Title: NOVEL GENES ENCODING A NYSTATIN POLYKETIDE SYNTHASE AND THEIR MANIPULATION AND UTILITY

Nystatin A1

(57) Abstract: The invention provides a nucleic acid molecule comprising: (a) a nucleotide sequence as shown in SEQ ID No. 35; or (b) a nucleotide sequence which is the complement of SEQ ID No. 35; or (c) a nucleotide sequence which is degenerate with SEQ ID No. 35; or (d) a nucleotide sequence hybridising under conditions of high stringency to SEQ ID No. 35, to the complement of SEQ ID No. 35, or to a hybridisation probe derived from SEQ ID No. 35 or the complement thereof; or (e) a nucleotide sequence having at least 80 % sequence identity with SEQ ID No. 35; or (f) a nucleotide sequence having at least 65 % sequence identity with SEQ ID No. 35 wherein said sequence preferably encodes or is complementary to a sequence encoding a nystatin PKS enzyme or a part thereof. Also provided are part of such molecules and polypeptides (and parts thereof) encoded by such a nucleic acid molecule, and the use of such molecules and polypeptides in facilitating nystatin biosynthesis and in the synthesis of nystatin derivatives and novel polyketide as macrolide structures.

WO 01/59126 A2

19, N-7024 Trondheim (NO). **GULLIKSEN, Ole-Martin** [NO/NO]; Grevlingvn. 31 D, N-0595 Oslo (NO).

(74) **Agents: DZIEGLEWSKA, Hanna** et al.; Frank B. Dehn & Co., 179 Queen Victoria Street, London EC4V 4EL (GB).

(81) **Designated States** *(national)*: AE, AG, AL, AM, AT, AT (utility model), AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, CZ (utility model), DE, DE (utility model), DK, DK (utility model), DM, DZ, EE, EE (utility model), ES, FI, FI (utility model), GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SK (utility model), SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) **Designated States** *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

GENE CLUSTER ENCODING A NYSTATIN POLYKETIDE
SYNTHASE AND ITS MANIPULATION AND UTILITY

5

        The present invention relates to the cloning and
sequencing of the gene cluster encoding a modular
polyketide synthase enzyme involved in the biosynthesis
of the macrolide antibiotic nystatin. The invention
10   thus relates to novel genes and nucleic acid molecules
encoding proteins/polypeptides exhibiting functional
activities involved in nystatin biosynthesis, such
functional proteins/polypeptides themselves, and their
uses both in facilitating nystatin biosynthesis and in
15   the synthesis of nystatin derivatives and novel
polyketide or macrolide structures.
        Polyketides are natural products synthesized by
microorganisms, many of which have applied potential as
pharmaceuticals or as agricultural or veterinary
20   products. Examples of polyketides used in medical
treatments include the antibiotics erythromycin
(antibacterial), nystatin (antifungal), avermectin
(antiparasitic), rapamycin (immunosuppressant) and
daunorubicin (antitumor). The Gram-positive bacteria
25   *Streptomyces* are the main producers of polyketides, and
the genetics and biochemistry of polyketide biosynthesis
in these organisms are relatively well characterized
(Hopwood *et al.*, Chem. Rev. v. 97: 2465-2497 (1997)).
Macrolide polyketide compounds are formed via repeated
30   condensations of simple carboxylic acids by modular
(type I) polyketide synthases (PKS) in a manner similar
to fatty acid biosynthesis. The modular hypothesis
proposed by Donadio *et al.* Science, v. 252: 675-679
(1991) suggested that type I PKSs are organized in
35   repeated units (modules), each of which is responsible
for one condensation cycle in the synthesis of a
polyketide chain. This was proven to be correct by

- 2 -

manipulations of PKSs type I genes resulting in
predictable changes in the chemical structures of
macrolides. Beside condensation of the next carboxylic
acid onto the growing polyketide chain, ensured by the
5   catalytic activity of the β-ketoacyl synthase (KS)
domain, modules of PKSs type I may contain domains with
β-ketoreductase (KR), dehydratase (DH) and enoyl
reductase (ER) activities, which determine the reduced
state of incorporated extender units. The
10   acyltransferase (AT) and acyl carrier protein (ACP)
domains present in each module are responsible for the
choice of extender unit and retention of the growing
polyketide chain on the PSK, respectively. Upon
completion of synthesis, the polyketide chain is
15   released from PKSs via action of a thioesterase (TE),
that is probably also involved in cyclization of the
final product. Thus, PKSs type I represent an assembly
line for polyketide biosynthesis, that can be
manipulated by changing the number of modules, their
20   specificities towards carboxylic acids, and by
inactivating or inserting domains with reductive
activities (Katz, Chem. Rev., v. 97, 2557-2575, 1997).
After the polyketide moiety of a macrolide is
synthesized and cyclized to form a macrolactone ring, it
25   is usually modified via hydroxylation, glycosylation,
methylation and/or acylation. These modifications are
believed to be crucially important for the biological
activities of macrolides.

        The genes for macrolide antibiotic biosynthesis in
30   *Streptomyces* are organized in clusters, and a number of
such clusters have already been identified.
Exploitation of recombinant DNA technology has made it
possible to isolate complete antibiotic biosynthetic
gene clusters by screening the gene libraries with DNA
35   probes encoding PKSs (Schwecke *et al.*, Proc. Natl. Acad
Sci. USA, v. 92: 7839-7843, 1995). The molecular
cloning and complete DNA sequencing has been described

- 3 -

for several macrolide antibiotic gene clusters of
Streptomyces, including those for avermectin, pikromycin
and rapamycin (Ikeda *et al.*, Proc. Natl. Acad. Sci. USA,
v. 96: 9509-9514, 1999; Xue *et al.*, Proc. Natl. Acad.
Sci. USA, v. 95: 1211-12116, 1998; Schwecke *et al.*,
Proc. Natl. Acad. Sci. USA, v. 92: 7839-7843, 1995).
Partial cloning and DNA sequencing of the gene cluster
for the polyene macrolide antibiotic pimaricin has
recently been reported (Aparicio *et al.*, J. Biol. Chem.,
v. 274: 10133-10139, 1999). However, a complete DNA
sequence of genes for the biosynthesis of a polyene
macrolide antibiotic with antifungal activity has not
yet been disclosed. There is a need and desire to
increase the repertoire of available antifungal
antibiotics, and/or to improve upon the properties (e.g.
efficacy, toxicity, etc.) of existing drugs. Hence the
provision of new antifungal treatments, particularly
those exhibiting new or improved properties would
represent a considerable advance in the art.

The present invention is directed to this aim, and
is based on the cloning and DNA sequencing of the
nystatin biosynthesis gene cluster. This provides the
first example of the identification of such antifungal
antibiotic biosynthesis genes, as well as a tool for
genetic manipulation in order to modify the properties
of nystatin and/or the producing organism, or to obtain
novel potentially useful compounds.

The polyene antifungal antibiotic nystatin A1, the
complete stereostructure of which (see Fig. 1) has been
determined by Lancelin & Beau, Tetrahedron Lett, v. 30:
4521-4524, (1989), is produced by *Streptomyces noursei*
ATCC11455. From the structure of the nystatin molecule,
which belongs to the class of macrolide compounds, we
predicted that its polyketide backbone is synthesized by
a PKS type I enzyme. Based on this assumption, and as
described in the Examples below, a genomic library of
Streptomyces noursei ATCC11455 was screened using a

- 4 -

specially designed probe obtained by PCR using primers
based on conserved amino acid sequences within known β-
ketoacyl synthase (KS) and acyl carrier protein (ACP)
domains of known modular PKS enzymes. This led to the
identification of a number of clones or fragments which
we have sequenced and shown to contain parts or portions
of the nystatin PKS gene cluster. We have further shown
that alteration of the fragment sequences to inactivate
the encoded product leads to abrogation of nystatin
biosynthesis (see Example 1 below), thereby confirming
the requirement of the identified PKS for nystatin
biosynthesis. Subsequent work on the clones/fragments
has lead to the sequencing of the PKS gene cluster, and
the identification of the different modules and
enzymatic domains, regulatory regions etc, within it.

Furthermore, as will be described in more detail
below, we have shown that manipulations of functional
DNA sequences within the novel nystatin PKS gene cluster
which we have identified, have led to the synthesis of
novel molecular structures, e.g. nystatin derivatives
with improved function. This opens up the exciting
possibility of manipulating the nystatin A1 PKS gene
cluster to obtain not only beneficial new nystatin
derivatives, but also to improve and facilitate the
biosynthetic production process (for example to improve
yield, or production conditions, or to expand the range
of available host cells) or to provide novel compounds
with new activities and/or properties.

More particularly, two primary regions (or "parts"
or "portions") of the nystatin PKS gene cluster were
initially identified and sequenced, together
representing approximately 80% of the nystatin PKS gene
cluster, and the functional sequences within said
regions (e.g. PKS genes, regulatory regions etc, as well
as functional gene products, enzymatic domains etc.)
have been identified and characterised.

The first region ("Region 1"), which we have termed

- 5 -

"Nys 1" and the complete DNA sequence of which is shown
in SEQ ID No. 1, has been shown to contain a number of
PKS or associated genes or regulatory regions, and 13
separate "features" or open reading frames (ORFs) have
5   been identified (the amino acid sequences of the
translation products of which are shown in SEQ ID Nos. 3
to 15 respectively).

The second region ("Region 2"), which we have
termed "Nys 2", and the complete coding sequence of
10   which is shown in SEQ ID No. 2, also comprises a number
of "functional" regions, and 5 separate "features" or
ORFs have been identified, the amino acid sequences of
the translation products of which are shown in SEQ ID
Nos. 16 to 20 respectively.

15   Nys 1 and Nys 2 (i.e. SEQ ID Nos. 1 and 2 and the
sequences they encode) are the subject of British Patent
Application No. 0002840.7 filed on 8 February 2000.

Subsequent sequencing efforts have led to the
determination of the sequence of the DNA spanning the
20   gap between SEQ ID Nos. 1 and 2, and the identification
of novel genes in this region. In addition, the partial
gene sequences contained in SEQ ID Nos. 1 and 2 encoding
the gene products NysI (SEQ ID No. 20) and NysDII (SEQ
ID Nos. 3) (see further below) have been completed (see
25   new SEQ ID Nos. 36 and 37 respectively - see further
below). Thus, these sequencing efforts have led to the
identification and sequencing of the DNA region
encompassing the entire nystatin PKS gene cluster, and
the identification and characterisation of the
30   functional sequences within this region.

The complete coding sequence for (i.e. the complete
nucleotide sequence encoding) the nystatin biosynthetic
gene cluster is shown in SEQ ID No. 35. This has been
shown to contain a number of PKS or associated genes or
35   regulatory regions, and 23 separate "features" or ORFs
have been identified (the amino acid sequences of the
translation products of which are shown in SEQ ID Nos. 2

- 6 -

to 19, and 36 to 42 respectively).

The complete coding sequence for (i.e. the complete nucleotide sequence encoding) the nystatin biosynthetic gene cluster, as shown in SEQ ID No. 35, is the subject of British Patent Application No. 0008786.6 filed on 10 April 2000 and British Patent Application No. 0009387.2 filed on 14 April 2000.

In one aspect, the present invention thus provides a nucleic acid molecule comprising:

(a)   a nucleotide sequence as shown in SEQ ID No. 35; or

(b)   a nucleotide sequence which is the complement of SEQ ID No. 35; or

(c)   a nucleotide sequence which is degenerate with SEQ ID No. 35; or

(d)   a nucleotide sequence hybridising under conditions of high stringency to SEQ ID No. 35, to the complement of SEQ ID No. 35, or to a hybridisation probe derived from SEQ ID No. 35 or the complement thereof; or

(e)   a nucleotide sequence having at least 80% sequence identity with SEQ ID No. 35; or

(f)   a nucleotide sequence having at least 65% sequence identity with SEQ ID No. 35 wherein said sequence preferably encodes or is complementary to a sequence encoding a nystatin PKS enzyme or a part thereof.

A "nystatin PKS enzyme" is defined further below, but briefly in the context of section (f) above means an enzyme or protein or polypeptide that is functional in the synthesis, transport or transfer of a macrolide antibiotic or polyketide moiety, preferably nystatin or a nystatin derivative or nystatin-related molecule.

In a further aspect, the present invention also provides a nucleic acid molecule comprising:

(a)   a nucleotide sequence as shown in SEQ ID No. 1 and/or in SEQ ID No. 2; or

(b)   a nucleotide sequence which is the complement of SEQ ID No. 1 and/or SEQ ID No. 2; or

(c)   a nucleotide sequence which is degenerate with SEQ

- 7 -

ID No. 1 and/or SEQ ID No. 2; or

(d)   a nucleotide sequence hybridising under conditions
of high stringency to SEQ ID No. 1 and/or SEQ ID No. 2,
to the complement of SEQ ID No. 1 and/or SEQ ID No. 2,
or to a hybridisation probe derived from SEQ ID Nos. 1
and/or 2 or the complements thereof; or

(e)   a nucleotide sequence having at least 65% sequence
identity with SEQ ID No. 1 and/or SEQ ID No. 2, wherein
said sequence preferably encodes or is complementary to
a sequence encoding a nystatin PKS enzyme or a part
thereof.

A nucleic acid molecule of the invention may be an
isolated nucleic acid molecule (in other words isolated
or separated from the components with which it is
normally found in nature) or it may be a recombinant or
a synthetic nucleic acid molecule.

The nucleic acid molecule of the invention encodes
(or comprises a nucleotide sequence encoding) the
nystatin A1 PKS enzyme, or a portion thereof e.g. a
sequence encoding a single domain, or comprises a
nucleotide sequence in the nystatin A1 PKS gene cluster
which is a functional or non-functional genetic element.
More precisely, the nucleic acid molecule of the
invention encodes one or more polypeptides, or comprises
one or more genetic elements having functional activity
in the synthesis of a macrolide antibiotic or a
polyketide moiety, preferably nystatin or a nystatin
derivative or nystatin-related molecule. Such
functional activity may be enzymatic activity e.g. an
activity involved in the synthesis or transport or
transfer of a polyketide moiety or a macrolide molecule
(this can be macrolide chain or ring synthesis or any
step contributory thereto, or macrolide ring or
polyketide chain modification etc) and/or it may be a
regulatory activity, e.g. regulation of the expression
of the genes or proteins involved in the synthesis, or
regulation of the synthetic process, and/or it may be a

- 8 -

"transporter activity". Thus, included generally are
also transport proteins involved in the transfer or
transport of polyketide or macrolide moieties e.g. in
the transport or efflux of the synthesised molecule
5   within or out of the cell. Also included in this
respect are glycosylation proteins which includes
molecules involved in the biosynthesis and/or attachment
of saccharides (e.g. mycosamine) to the macrolide or
polyketide.

10   Whilst nucleotide sequences encoding a desired
product are preferred according to the invention, also
encompassed are nucleotide sequences comprising
functional genetic elements such as promoters, promoter-
operator regions, enhancers, other regulatory sequences
15   etc. Thus, the nucleic acid molecule of the invention
need not comprise the entire PKS gene cluster but may
comprise a portion or part of it e.g. a part encoding a
polypeptide having a particular function or a regulatory
sequence. This may comprise one or more genes, and/or
20   regulatory sequences, and/or one or more modules or,
enzymatic domains, or non-coding or coding functional
genetic elements (e.g. elements controlling gene
expression, transcription, translation etc).

            In one such aspect, the invention provides a
25   nucleic acid molecule as defined above, wherein said
nucleotide sequence of SEQ ID No. 35 (or variant thereof
as defined in (b) to (f) above) does not include the
portions of the molecule comprising ORF 1 (see Table 1
below). In other words, in this embodiment, the
30   nucleotide sequence of SEQ ID No. 35 does not comprise
nucleotides 124026 to 125222.

            In another such aspect, the invention provides a
nucleic acid molecule as defined above, wherein said
nucleotide sequence of SEQ ID No. 35 (or variant thereof
35   as defined in (b) to (f) above) does not include the
portions of the molecule comprising ORF 2 (see Table 1
below). In other words, in this embodiment, the

- 9 -

nucleotide sequence of SEQ ID No. 35 does not comprise
nucleotides 122812 to 123876.

   In a further such aspect, the invention provides a
nucleic acid molecule as defined above, wherein said
5   nucleotide sequence of SEQ ID No. 35 (or variant thereof
as defined in (b) to (f) above) does not include the
portions of the molecule comprising NysF (see Table 1
below). In other words, in this embodiment, the
nucleotide sequence of SEQ ID No. 35 does not comprise
10   nucleotides 454 to 1191.

   Alternatively, the invention provides nucleic acid
molecules which contain a part of ORF 1, ORF 2 and/or
NysF, or a modified sequence of ORF 1, ORF 2 and/or
NysF, such that the expression or the function of the
15   ORF 1, ORF 2 and/or NysF gene product is ablated.

   Included within the scope of the invention are
nucleotide sequences which hybridise to SEQ ID Nos. 1 or
2 or 35 or their complements, or to parts thereof (i.e.
to hybridisation probes derived from SEQ ID Nos. 1 or 2
20   or 35 which are discussed in more detail below), under
high stringency conditions and which preferably encode
or are complementary to a sequence which encodes a
nystatin PKS enzyme or part thereof. Conditions of high
stringency may readily be determined according to
25   techniques well known in the art, as described for
example in Sambrook et al., 1989, Molecular Cloning, A
Laboratory Manual, 2nd Edition. Hybridising sequences
included within the scope of the invention are those
binding under non-stringent conditions (6 x SSC/50%
30   formamide at room temperature) and washed under
conditions of high stringency (e.g. 0.1 x SSC, 68°C),
where SSC = 0.15 M NaCl, 0.015M sodium citrate, pH 7.2.

   A hybridisation probe may be a part of the SEQ ID
No. 1 or SEQ ID No. 2 or SEQ ID No. 35 sequence (or
35   complementary sequence), which is of sufficient base
length and composition to function to hybridise to
sample or test nucleic acid sequences to determine

- 10 -

whether or not hybridisation under high stringency
condition occurs. The probe may thus be at least 15
bases in length preferably at least 30, 40, 50, 75, 100
or 200 bases in length. Representative probe lengths
thus include 30-500 bases e.g. 30-300, 50-200, 50-150,
75-100.

The hybridisation probe may be derived from a
coding or non-coding, functional or non-functional part
of the sequence (i.e. SEQ ID Nos. 1 or 2 or 35 or their
complements), and may for example correspond to a gene
or module or to an enzymatic domain, or a part thereof
(e.g. the part encoding the active site) or to a
sequence which links enzymatic domains or modules.
Thus, the hybridisation probe may have functional
activity in polyketide/macrolide synthesis as defined
above.

Nucleotide sequence identity may be determined
using the BestFit program of the Genetics Computer Group
(GCG) Version 10 Software package from the University of
Wisconsin. The program uses the local homology
algorithm of Smith and Waterman with the default values:
Gap creation penalty = 50, Gap extension penalty = 3,
Average match = 10,000, Average mismatch = -9.000.

Nucleotide sequences according to the invention may
exhibit at least 65%, 70%, 75%, 80%, 85%, 90%, 95% or
98% sequence identity with SEQ ID Nos. 1 or 2 or 35 and
preferably encode or are complementary to a sequence
which encodes a nystatin PKS enzyme or part thereof.
Nucleotide sequences meeting the % sequence identity
criteria defined herein may be regarded as
"substantially identical" sequences.

Where the nucleic acid molecules of the invention
are defined by reference to SEQ ID Nos. 1 and/or 2, the
nucleic acid molecule of the invention may thus comprise
a SEQ ID No. 1 or "SEQ ID No. 1-variant" sequence (i.e.
a sequence complementary, or degenerate to SEQ ID No. 1
or a functionally equivalent variant such as a

- 11 -

hybridising or substantially identical sequence as
defined above) or a SEQ ID No. 2 or "SEQ ID NO. 2-
variant" sequence or both. Nucleic acid molecules
comprising both a SEQ ID No. 1/SEQ ID NO. 1-variant
5   sequence and a SEQ ID No. 2/SEQ ID NO. 2-variant
sequence are preferred.

As referred to herein "functionally equivalent
variants" or "functional equivalents" retain at least
one function of the entity to which they are related (or
10   from which they are derived), e.g. encode a protein with
substantially the same properties, or exhibit
substantially the same regulatory or other functional
properties or activities.

As mentioned above, nucleic acid molecules
15   comprising parts or portions (e.g. fragments) of the
nucleotide sequences of SEQ ID No. 35 or of SEQ ID No. 1
and/or 2 (or their complementary, degenerate or
functionally equivalent variants) are also included
within the scope of the invention.

20   Such parts or portions of the PKS gene cluster
advantageously may also be regarded as functional
equivalents of the complete sequence. For example, the
sequence portion or fragment may retain a functional
activity as defined above, e.g. an enzymatic,
25   regulatory, or transporter activity in polyketide or
macrolide biosynthesis.

Conveniently, the part or portion of the PKS gene
cluster (e.g. of SEQ ID No. 35, or 1 and/or 2) is at
least 15 bases in length, more preferably at least 20,
30   25, 30, 35, 40, 50, 70, 100, 200, 300, 400, 500, 1000,
2000, 5000, 10,000, 15,000, 20,000, 30,000, or 50,000
bases. Representative fragment lengths thus include 15-
50,000 bases e.g. 50-30,000 bases, or 100-20,000, 100-
10,000 or 200-5,000, or 200-2,000. The part or portion
35   may comprise or encode contiguous or non-contiguous
nucleotide or amino acids.

Parts or portions of functional parts of the PKS

- 12 -

gene sequences are discussed in more detail below.

Parts or portions of the PKS gene cluster may also comprise the non-coding or non-functional part of the DNA molecule (or the nucleotide sequences), for example 5 promoter or operator sequences, or linker sequences joining individual genes, modules or enzymatic domains. These may be contiguous or non-contiguous and will be discussed further below.

As mentioned above, a number of genes and ORFS 10 within SEQ ID Nos. 35, 1 and 2 have been identified and such genes (or their complementary, degenerate or functionally equivalent variants as defined above) represent preferred "parts" or fragments of SEQ ID Nos. 35, 1 and 2. These are tabulated in Table 1 below:

15

Table 1

Molecule features of SEQ ID Nos. 35, 1 and 2 (the whole gene cluster sequence (125401 bp) Nys 1 (65140 bp) and 20 Nys 2 (27541bp) respectively)

SEQ ID No. 35			
Start	End	Gene	Description
1191	454 C	*nysF*	putative 4'-phsphopanthe-teine transferase
3092	1275 C	*nysG*	ABC transporter
4824	3070 C	*nysH*	ABC transporter
5122	6156	*nysDIII*	dGDP-mannose-4,6-dehydratase homolog
6338	34771	*nysI*	NysI PKS, modules 9-14
34792	51097	*nysJ*	NysJ PKS, modules 15-17
51155	57355	*nysK*	NysK PKS (module 18 + TE)
57503	58685	*nysL*	P450 monnoxygenase NysL
58980	58788 C	*nysM*	ferredoxin NysM

- 13 -

60241	59047 C	*nysN*	P450 monooxygenase NysN	
61296	60240 C	*nysDII*	putative aminotransferase	
62837	61317 C	*nysDI*	putative UDP-glucoronosyltransferase	
63067	67167	*nysA*	NysA PKS, loading module	
67213	76791	*nysB*	NysB PKS, modules 1 and 2	
76811	110101	*nysC*	NysC PKS, modules 3-8	
110521	111276	*nysE*	putative thioesterase	
111666	114566	*nysRI*	transcriptional activator	
114590	117451	*nysRII*	putative transcriptional activator	
117441	120224	*nysRIII*	putative transcriptional activator	
120676	121308.	*nysRIV* (short)	putative response regulator	
120628	121308	*nysRIV* (long)	putative response regulator	
121997	122758	*nysRV*	putative repressor	
123876	122812 C	ORF2	putative transcriptional regulator	
124026	125222	ORF1	putative peptidase (aminohydrolase)	

Note: "C" indicates that the gene is encoded by the complement DNA strand.

nys 1 (SEQ ID No. 1)

Start	End	Gene	Description Name
1035	3 C	*nysD2*	putative aminotransferase
2447	1058 C	*nysD1*	putative UDP-glucuronosyl-transferase
2806	6904	*nysA*	nystatin PKS, loading module
6952	16528	*nysB*	nystatin PKS, modules 1 and 2

- 14 -

16550	49838		*nysC*	nystatin PKS, modules 3-8
50227	51013		*nysE*	putative thioesterase
51405	54303		*nysR1*	putative transcriptional activator 1
54329	57188		*nysR2*	putative transcriptional activator 2
57180	59961		*nysR3*	putative transcriptional regulator 3
60367	61045		*nysR4*	putative response regulator
61736	62495		*nysR5*	putative repressor
63615	62553	C	ORF2	putative transcriptional regulator
63765	64959		ORF1	putative peptidase (aminohydrolase)

nys 2 (SEQ ID No. 2)

Start	End		Gene	Description Name
1191	456	C	*nysF*	putative 4'-phosphopantheteine transferase
3092	1277	C	*nysG*	putative ABC transporter
4824	3072	C	*nysH*	putative ABC transporter
5122	6154		*nysD3*	putative GDP-mannose-4,6-dehydratase
6338	27541		*nysI*	nystatin PKS, modules 9-13 (incomplete)

"C" in the table above refers to complementary strands.

It will be appreciated that *nys*D2, D1, A, B, C, E, R1 to R5, ORF1, ORF2 of SEQ ID No. 1 and *nys*F, G, H and D3, and the partial *nys*I sequences of SEQ ID No. 2, correspond to their named counterparts in SEQ ID No. 35, which represents the whole complete coding sequence for the gene cluster, and comprises the nucleotide sequences

- 15 -

of SEQ ID Nos. 1 and 2. There is however a difference
in the nucleotide numbering as between the corresponding
features in SEQ ID Nos. 2 and 35; for SEQ ID Nos. 1 and
2, the first nucleotide in the stop codon is recognised
as the end of the gene, whereas for SEQ ID No. 35, the
third nucleotide of the stop codon is recognised as the
end of the gene. *nys*D1, D2 and D3 are also known as
NysDI, DII and DIII, and *nys*R1 to R5 are also known as
*nys*RI to RV.

As regards SEQ ID No. 1 (*nys1*) and the gene
sequence *nysRI*, Table 1 shows this to comprise
nucleotides 51405 to 54303. Further sequence analysis
has revealed however that there are in fact two start
codons, nucleotides 51405-51407 encoding GTG and
nucleotides 51408-51410 encoding ATG. Accordingly,
nucleotide 51408 of SEQ ID No. 1 represents an
alternative start nucleotide for *nysRI*. ATG is
preferred as a start codon to GTG, and consequently
51408 is regarded the start nucleotide in future
references to *nysRI*. The start of *nysRI* in SEQ ID No.
35 is indicated in Table 1 above as the ATG codon, and
the translation product of *nysRI* shown in SEQ ID No. 9
below is deduced from nucleotides 51408-54303 of SEQ ID
No. 1. [In an alternative presentation of the NysRI
translation product, wherein nucleotide 51405 of SEQ ID
No. 1 is the start nucleotide, SEQ ID No. 9 is modified
by the inclusion of an additional "first" amino acid,
V].

As regards SEQ ID Nos. 35 and 1, and the gene
sequence *nysRIV*, Table 1 shows this to comprise
nucleotides 120676 to 121308 in SEQ ID No. 35 (nysRIV
short). Further sequence analysis has revealed however
that there are in fact two start codons, nucleotides
120676-120678 encoding GTG and nucleotides 120628-120630
encoding GTG. These start codons correspond to start
codons in SEQ ID No. 1 for nysR4 at nucleotides 60367-
60369 (as stated in Table 1) and 60415-60417. The

- 16 -

upstream GTG (120628-120630 in SEQ ID No. 35, which
corresponds to 60367-60369 in SEQ ID No. 1) is preferred
as a start codon (see Example 6), and consequently
120628 is regarded the start nucleotide in future

5   references to *nysRIV*. The start of *nysRIV* in SEQ ID No.
35 is indicated in Table 1 above (nysRIV short) as the
downstream start codon 120676-120678, and the deduced
translation product of *nysRIV* named herein as "NysRIV
short" is shown in SEQ ID No. 12. Table 1 also shows

10   the alternative start codon of *nysRIV* (*nysRIV* long) in
SEQ ID No. 35 as the upstream start codon 120628-120630.
Thus a preferred alternative presentation of the NysRIV
translation product, named herein "NysRIV (long)", is
shown in SEQ ID No. 43.

15   Alternative representative parts of the SEQ ID No.
35, or 1 and/or 2 sequences include the nucleotide
sequences between the respective "start" and "end"
nucleotide positions, either individually or
collectively.

20   The translation products of the respective "genes"
have been deduced and the amino acid sequences are set
out in the following SEQ ID Nos. shown below:

	Gene Product	SEQ ID No.
25	NysD2 (NysDII) (partial)	3
	NysD1 (nysDI)	4
	NysA	5
	NysB	6
	NysC	7
30	NysE	8
	NysR1 (NysRI)	9
	NysR2 (NysRII)	10
	NysR3 (NysRIII)	11
	NysR4 (NysRIV) (short)	12
35	NysR4 (NysRIV) (long)	43
	NysR5 (NysRV)	13
	ORF2	14

- 17 -

ORF1	15
NysF	16
NysG	17
NysH	18
NysD3(NysDIII)	19
NysI	20
NysDII(complete)	36
NysI(complete)	37
NysJ	38
NysK	39
NysL	40
NysM	41
NysN	42

SEQ ID Nos. 4, 10, 12, 13, 14, 16, 18, 41 and 43 show valine (V) as the first amino acid. According to practice in this field and conceptual translation of bacterial DNA, the first amino acid of a protein (translation product) is always methionine (M), regardless of the start codon. Accordingly, translation products and amino acid sequences of the present invention include not only SEQ ID Nos. 4, 10, 12, 13, 14, 16, 18, 41 and 43 as presented but also modifications of the aforesaid sequences in which the first V is replaced with M. References to SEQ ID Nos. 4, 10, 12, 13, 14, 16, 18, 41 and 43 below will be understood to include not only the sequences as presented, but also the said sequences wherein the first V is replaced with M.

Viewed from an alternative aspect, the present invention also provides a nucleic acid molecule comprising a nucleotide sequence encoding one or more amino acid sequences selected from SEQ ID Nos 3 to 20 or 36 to 43, or a nucleotide sequence which is complementary thereto or degenerate therewith.

Also provided are nucleic acid molecules comprising nucleotide sequences encoding one or more amino acid

- 18 -

sequences (i.e. polypeptides) which exhibit at least 60%
sequence identity with any one of SEQ ID Nos. 3 to 20 or
36 to 43.

5   A further aspect of the invention provides a
polypeptide encoded by a nucleic acid molecule of the
invention as defined herein.

More particularly, this aspect of the invention
provides a polypeptide comprising:

(a) all or part of an amino acid sequence as shown in
10   any one or more of SEQ ID Nos. 3 to 20 or 36 to 43; or
(b) all or part of an amino acid sequence which has at
least 60% sequence identity with any one or more of SEQ
ID Nos. 3 to 20 or 36 to 43.

In particular the amino acid sequence may exhibit
15   at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95% or 98%
identity with the polypeptide of any one of SEQ ID Nos.
3 to 20 or 36 to 43. Alternatively, the amino acid
sequence may exhibit at least 70%, 75%, 80%, 85%, 90%,
95% or 98% similarity with the polypeptide of any one of
20   SEQ ID Nos. 3 to 20 or 36 to 43. Amino acid
(polypeptide) sequences meeting the % sequence identity
or similarity criteria herein are regarded as
"substantially identical". The polypeptide of the
invention may be an isolated, purified or synthesized
25   polypeptide. The term "polypeptide" is used herein to
include any amino acid sequence of two or more amino
acids i.e. both short peptides and longer lengths (i.e.
polypeptides) are included.

Amino acid sequence identity or similarity may be
30   determined using the BestFit program of the Genetics
Computer Group (GCG) Version 10 Software package from
the University of Wisconsin. The program uses the local
homology algorithm of Smith and Waterman with the
default values: Gap creation penalty = 8, Gap extension
35   penalty = 2, Average match = 2.912, Average mismatch = -
2.003.

A "part" of the amino acid sequence of any one of

- 19 -

SEQ ID Nos. 3 to 20 or 36 to 43 (or of a "substantially identical" sequence as defined above) may comprise at least 20 contiguous amino acids, preferably at least 30, 40, 50, 70, 100, 150, 200, 300, 400, 500, 1,000, 2,000, 5,000 or 10,000 contiguous amino acids.

The polypeptide, and preferably also the part thereof, is functionally active according to the definitions given above, e.g. is enzymatically active or has a regulatory or transport functional activity. The part may not itself be functionally active but may in some instances provide regions with functional properties of the whole, e.g. represent the active site or co-factor binding site required for enzymatic activity.

The studies described in the Examples below have characterised the nucleotide and polypeptide sequences of the invention and various functional regions within them have been identified. Such functional regions form separate aspects of the invention. For the various translation products (i.e. gene products of SEQ ID Nos. 3 to 20 or 36 to 43), these functional regions are summarised below in Table 2 below:

Table 2: Molecule Features of Translation Products of
         SEQ ID Nos. 3 to 20 and 36 to 43

(i)   SEQ ID No.: 5
      Translation Product Name : NysA

Start AA	End AA	Name	Description
8	430	KS^s	KS domain, loading module
528	841	AT	AT domain, loading module
855	1055	DH	DH domain, loading module
1285	1359	ACP	ACP domain, loading module

- 20 -

(ii) SEQ ID No.: 6
     Translation Product Name: NysB

Start	End	Name	Description
42	462	KS1	KS domain, module 1
578	897	AT1	AT domain, module 1 (mMCoA-specific)
911	1110	DH1	DH domain, module 1 (inactive)
1201	1447	KR1	KR domain, module 1
1484	1559	ACP1	ACP domain, module 1
1579	2004	KS2	KS domain, module 2
2117	2439	AT2	AT domain, module 2 (mMCoA-specific)
2453	2659	DH2	DH domain, module 2 (inactive)
2749	2996	KR2	KR domain, module 2
3025	3102	ACP2	ACP domain, module 2

(iii) SEQ ID No.: 7
      Translation Product Name: NysC

Start	End	Name	Description
35	455	KS3	KS domain, module 3
546	858	AT3	AT domain, module 3
872	1073	DH3	DH domain, module 3
1381	1628	KR3	KR domain, module 3
1662	1735	ACP3	ACP domain, module 3
1757	2180	KS4	KS domain, module 4
2291	2603	AT4	AT domain, module 4
2617	2818	DH4	DH domain, module 4
3124	3371	KR4	KR domain, module 4
3407	3480	ACP4	ACP domain, module 4
3501	3924	KS5	KS domain, module 5

- 21 -

4032	4346	AT5	AT domain, module 5
4360	4561	DH5	DH domain, module 5
4953	5239	ER5	ER domain, module 5
5248	5495	KR5	KR domain, module 5
5528	5601	ACP5	ACP domain, module 5
5623	6046	KS6	KS domain, module 6
6166	6478	AT6	AT domain, module 6
6492	6704	DH6	DH domain, module 6
7038	7281	KR6	KR domain, module 6
7315	7388	ACP6	ACP domain, module 6
7408	7831	KS7	KS domain, module 7
7939	8253	AT7	AT domain, module 7
8267	8470	DH7	DH domain, module 7
8812	9086	KR7	KR domain, module 7
9120	9193	ACP7	ACP domain, module 7
9214	9637	KS8	KS domain, module 8
9758	10072	AT8	AT domain, module 8
10086	10289	DH8	DH domain, module 8
10657	10904	KR8	KR domain, module 8
10939	11012	ACP8	ACP domain, module 8

(iv) SEQ ID No.: 9
      Translation Product Name: Nys R1

Start	End	Name	Description
42	49	P-loop	ATP/GTP binding site motif A
904	932	HTH	LuxR-type helix-turn-helix motif (DNA binding)

(N.B. In the alternative representation of NysR1 above, where nt 51405 of SEQ ID No. 1 is regarded as the start codon, these start and amino acid end numbers would each increase by 1.)

- 22 -

Start	End	Name	Description
(v)	SEQ ID No.: 10		
	Translation Product Name: NysR2		
902	930	HTH	LuxR-type helix-turn-helix motif (DNA binding)

Start	End	Name	Description
(vi)	SEQ ID No.: 11		
	Translation Product Name: NysR3		
26	47	LZ	Leucine zipper motif (DNA binding)
548	568	TM1	Transmembrane domain (putative)
583	610	TM2	Transmembrane domain (putative)
884	912	HTH	LuxR helix-turn-helix motif

Start	End	Name	Description
(vii)	SEQ ID No.: 12		
	Translation Product Name: NysR4 (short)		
97	104	P-loop	ATP/GTP binding site motif A
149	177	HTH	LuxR helix-turn-helix motif (DNA binding)

- 23 -

(viii) SEQ ID No.: 13
        Translation Product Name: NysR5

Start End   Name   Description

  6   40   HTH   DeoR helix-turn-helix motif (DNA
                      binding)

(ix) SEQ ID No.: 14
       Translation Product Name: ORF2

Start End   Name   Description

 186   202   HTH   AsnC HTH motif signature

(x)   SEQ ID No.: 17
       Translation Product Name: NysG

Start   End   Name   Description

  31   313   TM   Transmembrane regions
 392   399   P-loop   ATP/GTP binding site
 496   510   ABC   ABC transporters signature

(xi) SEQ ID No.: 18
       Translation Product Name: NysH

Start   End   Name   Description

  17   288   TM   Transmembrane regions
 368   375   P-loop   ATP/GTP binding motif A
 472   486   ABC   ABC transporters signature

- 24 -

(xii) SEQ ID No.: 20
       Translation Product Name: NysI (partial)

Start	End	Name	Description
34	448	KS9	KS domain, module 9
572	890	AT9	AT domain, module 9
904	1123	DH9	DH domain, module 9
1443	1686	KR9	KR domain, module 9
1720	1793	ACP9	ACP domain, module 9
1813	2236	KS10	KS domain, module 10
2346	2664	AT10	AT domain, module 10
2678	2890	DH10	DH domain (inactive), module 10
2983	3229	KR10	KR domain, module 10
3266	3339	ACP10	ACP domain, module 10
3358	3780	KS11	KS domain, module 11
3898	4217	AT11	AT domain, module 11 (mMCoA-specific)
4231	4432	DH11	DH domain (inactive), module 11
4523	4770	KR11	KR domain, module 11
4806	4879	ACP11	ACP domain, module 11
4801	5325	KS12	KS domain, module 12
5432	5754	AT12	AT domain, module 12
5768	5977	DH12	DH domain (inactive), module 12
6068	6315	KR12	KR domain, module 12
6348	6421	ACP12	ACP domain, module 12
6454	6873	KS13	KS domain, module 13

(xiii) SEQ ID No.: 37
        Translation Product Name: NysI

Start	End	Name	Description
34	448	KS9	KS domain, module 9
572	890	AT9	AT domain, module 9
904	1123	DH9	DH domain, module 9

- 25 -

	1443	1686	KR9	KR domain, module 9
	1720	1793	ACP9	ACP domain, module 9
	1813	2236	KS10	KS domain, module 10
	2346	2664	AT10	AT domain, module 10
5	2678	2890	DH10	inactive DH domain, module 10
	2983	3229	KR10	KR domain, module 10
	3266	3336	ACP10	ACP domain, module 10
	3355	3777	KS11	KS domain, module 11
	3898	4217	AT11	AT domain (methylmalony-CoA-
10				specific), module 11
	4231	4432	DH11	inactive DH domain, module 11
	4523	4769	KR11	KR domain, module 11
	4806	4879	ACP11	ACP domain, module 11
	4901	5325	KS12	KS domain, module 12
15	5432	5754	AT12	AT domain, module 12
	5768	5977	DH12	inactive DH domain, module 12
	6068	6315	KR12	KR domain, module 12
	6348	6421	ACP12	ACP domain, module 12
	6454	6873	KS13	KS domain, module 13
20	6973	7293	AT13	AT domain, module 13
	7307	7448	DH13	inactive DH domain, module 13
	7535	7774	KR13	inactive KR domain, module 13
	7813	7886	ACP13	ACP domain, module 13
	7908	8323	KS14	KS domain, module 14
25	8430	8741	AT14	AT domain, module 14
	8755	8962	DH14	inactive DH domain, module 14
	9050	9296	KR14	KR domain, module 14
	9319	9394	ACP14	ACP domain, module 14

30 _____

(xiii) SEQ ID No.: 38
    Translation Product Name: NysJ

_____

Start   End   Name   Description

35

	41	464	KS15	KS domain, module 15
	578	889	AT15	AT domain, module 15

- 26 -

903	1102	DH15	DH domain, module 15
1446	1731	ER15	ER domain, module 15
1740	1988	KR15	KR domain, module 15
2023	2096	ACP15	ACP domain, module 15
2117	2538	KS16	KS domain, module 16
2635	2953	AT16	AT domain, module 16
2967	3167	DH16	inactive DH domain, module 16
3257	3500	KR16	KR domain, module 16
3539	3612	ACP16	ACP domain, module 16
3634	4057	KS17	KS domain, module 17
4153	4472	AT17	AT domain, module 17
4486	4725	DH17	inactive DH domain, module 17
4997	5245	KR17	KR domain, module 17
5277	5350	ACP17	ACP domain, module 17

(xiii) SEQ ID No.: 39
     Translation Product Name: NysK

Start	End	Name	Description
34	457	KS18	KS domain, module 18
568	881	AT18	AT domain, module 18
898	1102	DH18	inactive DH domain, module 18
1416	1663	KR18	KR domain, module 18
1695	1769	ACP18	ACP domain, module 18
1849	2066	TE	thioesterase domain

(xiv) SEQ ID No.: 43
     Translation Product Name: NysRIV (long)

Start	End	Name	Description
31	85	PAS	PAS-like domain
113	120	P-loop	ATP/GTP binding site motif A
165	193	HTH	LuxR helix-turn-helix motif (DNA binding)

- 27 -

As referred to in the above "inactive" denotes DH
domains which lack the conserved amino acid sequence
representing the active site motif $H(X_3)G(X_4)P$ found in
DH in other PKSs. It will however be appreciated that
5   these domains may have activity although this is likely
to be distinct from the activity of DH domains in other
PKSs.

It will be seen that SEQ ID Nos. 5 (NysA), 6
(NysB), 7 (NysC), 20 and 37 (NysI), 38 (NysJ) and 39
10   (NysK) constitute actual "PKS" enzymes, namely enzymes
involved in polyketide synthesis. These gene products
contain identifiable enzymatic domains and modules which
are tabulated in Table 2 above and shown also in Figures
4, 8 and 9 (see also Example 1 and Table 4, and Example
15   4 below which describes the DNA sequence analysis of
nystatin biosynthesis gene cluster in more detail).
Such individual domains and molecules, as identified
herein form separate aspects of the present invention.

SEQ ID NOs 3 and 26 (NysDII), 4 (NysDI), 8 (NysE),
20   16 (NysF), 19 (NysDIII), 40 (NysL), 41 (NysM) and 42
(NysN) represent other enzymes functional in polyketide
or macrolide synthesis e.g. in polyketide release from
PKS, post-translational PKS modification, and polyketide
modification. SEQ ID NOs 10 to 15 and 43 (NysRI to
25   NysRV, and ORF2) respectively represent transcriptional
regulators, and SEQ ID NOs 17 and 18 (NysG and NysH)
represent transport proteins which are presumed to be
involved in polyketide transport from the cell. This is
also described in more detail in the Examples below.
30   Such functional proteins represent separate aspects of
the present invention. Also included are functional
parts or fragments of such proteins i.e. active parts or
fragments which retain (i.e. exhibit measurable levels
of) the biological activity of the parent molecule from
35   which they are derived (i.e. of the whole protein or
polypeptide).

The nucleotide sequences of the present invention

- 28 -

provide important tools and information which can be
utilised in a number of ways to manipulate nystatin
biosynthesis, to synthesise new nystatin derivatives or
novel polyketide or macrolide structures, and to provide
5   novel or modified PKS systems (by "PKS system" here is
meant a polyketide synthesis system i.e. a gene cluster
or protein complex, collection or assembly, which is
functional in polyketide synthesis, but which is not
necessarily restricted to PKS enzymes or enzymatic
10   domains, and which may contain also other functional
activities, e.g. other enzymatic (e.g. modificatory) or
transporter or regulatory functional proteins).

Thus, for example, the entire nystatin PKS gene
cluster or PKS synthetic system as provided herein, or a
15   portion thereof, may be subjected to modification so as
to modify one or more genes, or one or more modules, or
enzymatic domains, or functional sequences within it.
Such modified or derivatised PKS systems may be used to
synthesize novel or modified polyketide moieties, as
20   will be described in more detail below. In this
situation, the nystatin PKS system provided herein, or a
fragment or portion thereof, may function as an "origin"
or "template" or "source" system or sequence for
modification.

25   More particularly, in one such embodiment and as
further described below, the non-functional parts (e.g.
non-biologically active parts) of said system may be
utilised as a "scaffold", and the functional parts (e.g.
sequences encoding enzymatic portions) may be modified
30   to yield the derivative or modified PKS system. In some
embodiments only a single selected, or few selected
functional (e.g. enzymatic) regions may be modified,
leaving the remaining sequence or structure largely
intact.

35   Alternatively, the functional portions may be
utilised as tools or materials for the modification of
other "scaffold" structures e.g. individual nystatin

genes, modules or domains may be used for introduction
(e.g. insertion or replacement) into other PKS scaffold
structures e.g. PKS scaffold systems derived from PKS
systems for other macrolide antibiotics e.g.
5   erythromycin, rapamycin etc.
Included within the scope of the invention are
synthetic or recombinant polyketide synthase enzymes
derived from the scaffold encoded by the nystatin gene
cluster which are modified to include one or more
10   functional units derived from other modular enzymes.
Such functional units may encode a catalytic or
transport protein domain for example a ketoreductase
domain from a PKS enzyme or an ACP domain from a modular
hybrid polyketide/peptide synthesising enzyme. Such
15   domains can be derived from enzyme domain DNA sequences
from, for example, polyketide synthesising enzymes,
peptide synthesising enzymes, hybrid peptide polyketide
synthesising enzymes, fatty acid synthesising enzymes or
other enzyme domains known in the art. Analogously,
20   there are included within the scope of the invention,
synthetic or recombinant polyketide synthase enzymes
derived from the scaffold encoded by a different
polyketide synthase gene cluster, or modular enzyme
encoding gene cluster, which are modified to include one
25   or more functional units derived from the nystatin gene
cluster.
Thus, the sequence and activity information
provided here for the nystatin biosynthesis gene cluster
may be used to alter existing known gene clusters and
30   hence the products they produce. In particular,
selection and incorporation of particular domains
described herein (or modification of existing sequences)
into existing PKS gene clusters will allow incorporation
of particular properties attributable to the nystatin
35   gene cluster.
Thus, in a very general sense, the present
invention provides the use of the nucleic acid molecules

- 30 -

of the invention as defined herein in the preparation of a modified PKS system, or in the preparation of modified polyketide molecules.

Such novel or modified polyketide or macrolide molecules form a separate aspect of the present invention.

The nucleotide sequences may be utilised in this way according to the present invention in a random or directed or designed manner, e.g. to obtain and test a particular predetermined or pre-designed structure, or to create random molecules, for example libraries of polyketide structures, e.g. for screening (this is also described in more detail below).

Whether for modification within the nystatin-PKS scaffold, or for introduction into an alternative scaffold structure, the genes or genetic elements which can be modified include not only the actual PKS genes (which encode NysA, NysB, NysC, NysI, NysJ and NysK) or the individual molecules or domains thereof, but also genes encoding other enzymes or functional proteins involved in nystatin biosynthesis and transport (referred to herein collectively as "PKS genes" or "nystatin genes").

As regards the actual PKS genes, as will be described in more detail below, these may be modified to change the nature of the loading domain molecule which determines the nature of the starter unit, the number of modules, the nature of the extender, as well as the various dehydratase, reductase and synthase activities which determine the structure of the polyketide chain.

Other genes which can be modified include the thioesterase gene, (encoding NysE; SEQ ID NO. 8), which may be modified to increase the efficiency of the PKS system (in the case of a thioesterase having "editing" activity which clears the inappropriate substrates from the PKS). If the thioesterase simply cleaves the final product off the PKS, it can be used for making nystatin

- 31 -

derivatives with a smaller macrolactone ring by
truncating nystatin PKS, and fusing this thioesterase to
the end of the truncated protein via genetic
engineering.

5   Regulatory genes: activators can be overexpressed
and repressors inactivated in order to boost polyketide
or antibiotic production. This may be of particular
importance for the production of new nystatin
derivatives in recombinant strains, (which may be

10   produced in very small quantities).

The putative 4'-phosphopantetheine (PPT)
transferase gene (encoding NysF; SEQ ID NO: 16) can be
overexpressed in order to achieve efficient post-
translational modification and full functionality of the

15   PKS. It can also be used for expression of the nystatin
(or other) PKS in a heterologous host, which lacks the
specific PPT activity. Such hosts may include E.coli,
Saccharomyces cereviseae, etc.

Deoxysugar genes: glycosyltransferase (encoding

20   NysDI; SEQ ID NO:4) can be overexpressed in order to
boost glycosylation of the synthesised molecules e.g.
novel nystatin derivatives. It can also be modified by
in vitro mutagenesis in order to increase its specifi-
city towards the new substrates. Inactivation of this

25   glycosylstranferase will result in a recombinant strain
producing non-glycosylated nystatin (probably also
lacking some modifications) which can be used, for
example, for chemical modifications, or enzymatic assays
for screening new modification activities.

30   Aminotransferase (NysDII; SEQ ID Nos: 3 and 36) may
be inactivated to give a nystatin derivative. This
enzyme is presumed to attach the amino group on the
deoxysugar mycosamine. This gene may also be expressed
in other streptomycetes in order to achieve the same

35   reaction with another deoxysugar normally lacking an
amino group.

ABC transporters (e.g. NysH and NysG; SEQ ID NOs 17

- 32 -

and 18): can be overexpressed in order to make the
efflux of nystatin and its derivatives more efficient.
They may also be mutated in order to shift their
specificity towards different compounds. They may be
5   inactivated, if it is desired for any reason to
accumulate the nystatin or its derivatives inside the
cell.

The genes encoding monooxygenases NysL and NysN
(SEQ ID Nos. 40 and 42) can be inactivated in *S. noursei*
10   in order to obtain non-hydroxylated and non-oxidized
nystatin derivatives. Alternatively, they can be
mutated with the aim of changing their specificities
toward nystatin precursors. Overexpression of *nysL* and
nysN may potentially lead to increased yield of nystatin
15   or its derivatives if the hydroxylation and/or oxidation
steps are limiting in the nystatin biosynthetic pathway.
Genetic manipulations with *nysM* encoding ferredoxin (SEQ
ID No.41) might also be useful if one wishes to
establish an *in vitro* P450 hydroxylase system for
20   modifications of nystatin precursors.

Thus, in addition to modification of the nystatin
PKS system, or modification of other PKS systems by
using the "nystatin" genes, the nucleic acid molecules
of the invention can also be utilised to manipulate or
25   facilitate the biosynthetic process, for example by
extending the host range or increasing yield or
production efficiency etc.

In order to enable practice of the invention
according to the principles above, the invention also
30   provides an expression vector, and host cells containing
a nucleic acid molecule as herein defined.

Also provided are methods for production of a
polyketide or macrolide molecule (e.g. nystatin or a
nystatin derivative), comprising expressing within a
35   host cell, a nucleic acid molecule as defined above.
The polyketide or macrolide molecule produced within the
host cell or secreted or exported by the host-cell as a

- 33 -

result of expression of the nucleic acid molecule (i.e. expression of the introduced "PKS synthesis machinery") may then be recovered.

This method of the invention may thus involve growing or cultivating the host cell under conditions whereby the nucleic acid molecule is expressed, and allowing the expression product(s) of the nucleic acid molecule to synthesise the polyketide/macrolide molecule, or in other words, growing or cultivating the host cell under conditions wherein the polyketide or macrolide is produced.

Also provided are methods for preparing recombinant nucleic acid molecules according to the invention, comprising inserting the nucleic acid molecules containing the nucleotide sequences of the invention into another nucleic acid molecule, e.g. into vector nucleic acid, e.g. vector DNA.

Expression vectors of the invention may include appropriate control sequences such as for example translational (e.g. start and stop codons, ribosomal binding sites) and transcriptional control elements (e.g. promoter-operator regions, termination stop sequences) linked in matching reading frame with the nucleic acid molecules of the invention.

Vectors according to the invention may include plasmids and viruses (including both bacteriophage and eukaryotic viruses) according to techniques well known and documented in the art, and may be expressed in a variety of different expression systems, also well known and documented in the art.

A variety of techniques are known and may be used to introduce such vectors into prokaryotic or eukaryotic cells for expression, or into germ line or somatic cells to form transgenic animals. Suitable transformation or transfection techniques are well described in the literature.

The invention also includes transformed or

- 34 -

transfected prokaryotic or eukaryotic host cells, for
transgenic organisms containing a nucleic acid molecule
according to the invention as defined above. Such host
cells may for example include prokaryotic cells such as
5   E.coli, Streptomyces and other bacteria, eukaryotic
cells such as yeasts or the baculovirus-insect cell
system, transformed mammalian cells and transgenic
animals and plants.
        The nucleic acid molecules contained in the
10   expression vectors, and host cells and organisms etc.
above may also be, as will be described in more detail
below, derivative nucleic acid molecules, derived from
the nucleic acid molecules defined above, either by
modification or by introducing said molecules or parts
15   thereof into, or combining with, other nucleic acid
molecules.
        Thus, in one aspect, the invention provides
recombinant materials for the production of
combinatorial libraries of polyketides wherein the
20   polyketide members of the library are synthesized by
modified PKS systems derived from the naturally
occurring nystatin Al system provided herein by using
this system as a scaffold. Generally, many members of
these libraries may themselves be novel compounds, and
25   the invention further includes novel polyketide members
of these libraries. The invention methods may thus be
directed to the preparation of an individual polyketide.
The polyketide may or may not be novel, but the method
of preparation permits a more convenient method of
30   preparing it. The resulting polyketides may be further
modified to convert them to antibiotics, typically
through glycosylation. The invention also includes
methods to recover novel polyketides with desired
binding activities by screening the libraries of the
35   invention.
        Thus, in one aspect, the invention is directed to a
method of preparing a nucleic acid molecule which

contains or comprises a nucleotide sequence encoding a
modified polyketide synthase enzyme or enzyme system.
The method comprises using the nystatin PKS encoding
sequence as provided herein as a scaffold and modifying
5   the portions of the nucleotide sequence that encode
enzymatic or other functional activities e.g. by
mutagenesis, inactivation (e.g. by deletion or
insertion), or replacement. The thus modified
nucleotide sequence encoding a modified PKS can then be
10   used to modify a suitable host cell and the cell thus
modified employed to produce a polyketide different from
that produced by the native nystatin PKS, whose
scaffolding has been used to support modifications of
enzymatic activity.

15   Alternatively, one or more portions of the
nucleotide sequence that encode enzymatic or other
functional activities may be introduced into an
alternative (i.e. different "second") PKS scaffold (i.e.
a scaffold derived from a further "second" PKS system,
20   different from the nystatin PKS system).

         The invention is also directed to polyketides thus
produced and the antibiotics to which they may then be
converted.

         In another aspect, the invention is directed to a
25   multiplicity of cell colonies comprising a library of
colonies wherein each colony of the library contains an
expression vector for the production of a different
modular PKS, but derived from the nystatin PKS of the
invention, as defined above. The library of different
30   modular PKS may be obtained by modifying one or more of
the regions of a naturally occurring "nystatin" gene or
gene cluster encoding an enzymatic activity so as to
alter that activity, leaving intact the scaffold
portions of the naturally occurring gene.

35   In another aspect, the invention is directed to a
multiplicity of cell colonies comprising a library of
colonies wherein each colony of the library contains a

- 36 -

different modular PKS derived from the nystatin PKS of
the invention. The invention is also directed to
methods to produce libraries of PKS complexes and to
produce libraries of polyketides by culturing these
colonies, as well as to the libraries so produced. In
addition, the invention is directed to methods to screen
the resulting polyketide libraries and to novel
polyketides contained therein.

As mentioned above, a structural and functional
sequence analysis of the nystatin PKS gene
cluster/system is presented in the Examples below and
the DNA sequences are shown in SEQ ID NO 1, 2 and 35,
and further analysed in Tables 1 and 2 above, SEQ ID NOs
3 to 20, and 36 to 42 and in Figures 2, 3, 7, 8 and 9.
The modular and "domain" encoding structure of the "PKS"
gene may be seen. A module may typically contain a
ketosynthase (KS), an acyltransferase (AT) and an acyl
carrier protein (ACP). These three functions are
sufficient to activate an extender unit and attach it to
the remainder of the growing molecule. Additional
activities that may be included in a module relate to
reactions other than the Claisen condensation, and
include a dehydratase activity (DH), an enoylreductase
activity (ER) and a ketoreductase activity (KR). The
loading module catalyses the initial condensation, i.e.
it begins with a "loading domain" represented by AT and
ACP, which determine the nature of the starter unit.
The "finishing" of the molecule is believed to be
regulated by thioesterase activity (TE) and it is
believed that this is achieved by the TE activity
embedded in NysK. This thioesterase appears to catalyse
cyclization of the macrolide ring thereby increasing the
yield of the polyketide product. The NysE TE activity
is believed to be an "editing" one, and participates in
cleaving off certain substrates from the nystatin PKS
complex.

It will be seen from the sequences, Figures and

- 37 -

Tables above and below, that the regions in the genes
and modules that encode enzymatic activities are
separated by linker or "scaffold"-encoding regions.
These scaffold regions encode amino acid sequences that
5   space the enzymatic/functional activities at the
appropriate distances and in the correct order. Thus,
these linker regions collectively can be considered to
encode a "scaffold" into which the various activities
are placed in a particular order and spatial
10   arrangement. It should however be noted that in some
instances regions of the scaffold may be deleted or
modified without significantly affecting the activity of
the resultant PKS. Indeed the sequence encoding some
domains with functional activity may be fully or
15   partially deleted without significant effects. Thus as
used herein "scaffold" refers to portions of PKS cluster
not directly attributable to functional e.g. enzymatic
activities of said PKS, but responsible for maintenance
of its overall activity, e.g. providing correct spatial
20   orientation or structure. Said scaffold may comprise
all linker and non-functional regions of the PKS or a
functionally active (i.e. retaining structural
integrity) part thereof. This organization is similar
in other naturally occurring modular PKS gene clusters.
25   The invention provides libraries or individual
modified forms, ultimately of polyketides, by generating
modifications in the nystatin PKS gene cluster so that
the protein complexes produced by the cluster have
altered activities in one or more respects, and thus
30   produce polyketides other than the natural product of
the PKS (i.e. nystatin A1). Novel polyketides may thus
be prepared, or polyketides in general prepared more
readily, using this method. By providing a large number
of different genes or gene clusters derived from the
35   naturally occurring nystatin PKS gene cluster, each of
which has been modified in a different way from the
native cluster, an effectively combinatorial library of

- 38 -

polyketides can be produced as a result of the multiple
variations in these activities. Alternatively the
nystatin PKS "functional regions" (e.g. genes, modules
or domains) may be used, for introduction into a
5   "scaffold" obtained from another naturally occurring PKS
system. The modified PKS encoding sequences and systems
used in the present invention thus represent modular
polyketide synthases "derived from" a naturally
occurring nystatin PKS.
10   By a modular PKS "derived from" the nystatin PKS is
meant a modular polyketide synthase (or its
corresponding encoding gene(s)) that retains the
scaffolding of all of the utilized portion of the
naturally occurring gene. (Not all modules or genes
15   need be included in the constructs). On the constant
scaffold, at least one enzymatic or functional activity
is mutated, deleted or replaced, so as to alter the
activity. Alteration results when these activities are
deleted or are replaced by a different version of the
20   activity, or simply mutated in such a way that a
polyketide other than the natural product results from
these collective activities. This occurs because there
has been a resulting alteration of the starter unit
and/or extender unit, and/or stereochemistry, and/or
25   chain length or cyclization and/or reductive or
dehydration cycle outcome at a corresponding position in
the product polyketide. Where a deleted activity is
replaced, the origin of the replacement activity may
come from a corresponding activity in a different
30   naturally occurring polyketide synthase or from a
different region of the same PKS. Alternatively, such a
"derived" modular PKS may incorporate one or more
enzymatic or other functional activities (or their
encoding nucleotide sequences) obtained or derived from
35   the nystatin PKS described herein, in the scaffolding of
a second, different modular PKS (or its gene).
        Modification or manipulation of the modular PKS may

involve truncation, e.g. gene or domain or module
deletion or domain/gene/module swapping, addition or
inactivation, which may involve insertion or deletion.
Alternatively, random or directed modifications (i.e.
5   mutations) may be made in the nucleotide sequence of the
selected portion (e.g. in a gene/domain/module etc).

The derivative may contain preferably at least a
thioesterase activity from the nystatin PKS gene
cluster.

10   Advantageously, a polyketide synthase "derived
from" the nystatin PKS may contain the scaffolding
encoded by all or the portion employed of the nystatin
synthase gene, contains at least two modules that are
functional, preferably four or more modules and contains
15   mutations, deletions, or replacements of one or more of
the activities of these functional modules so that the
nature of the resulting polyketide is altered. This
definition applies both at the protein and genetic
levels. Particular preferred embodiments include those
20   wherein a KS, AT, KR, DH or ER has been inactivated or
deleted or replaced by a version of the activity from a
different PKS or from another location within the same
PKS. Also preferred are derivatives where at least one
noncondensation cycle enzymatic activity (KR, DH or ER)
25   has been deleted or wherein any of these activities has
been mutated so as to change the ultimate polyketide
synthesized.

Thus, there are five degrees of freedom for
constructing a polyketide synthase in terms of the
30   polyketide that will be produced. First, the polyketide
chain length will be determined by the number of modules
in the PKS system. Second, the nature of the carbon
skeleton of the PKS will be determined by the
specificities of the acyl transferases which determine
35   the nature of the extender units at each position --
e.g. malonyl, methyl malonyl, or ethyl malonyl, etc.
Third, the loading domain specificity will also have an

- 40 -

effect on the resulting carbon skeleton of the
polyketide. Thus, the loading domain may use a
different starter unit, such as acetyl, propionyl, and
the like. Fourth, the oxidation state at various
5   positions of the polyketide will be determined by the
dehydratase and reductase portions of the modules. This
will determine the presence and location of ketone,
alcohol, double bonds or single bonds in the polyketide.
     Finally, the stereochemistry of the resulting
10   polyketide is a function of three aspects of the
synthase. The first aspect is related to the AT/KS
specificity associated with substituted malonyls as
extender units, which affects stereochemistry only when
the reductive cycle is missing or when it contains only
15   a ketoreductase since the dehydratase would abolish
chirality. Second, the specificity of the ketoreductase
will determine the chirality of any β-OH. Finally, the
enoyl reductase specificity for substituted malonyls as
extender units will influence the result when there is a
20   complete KR/DH/ER available.
     Thus, the modular nystatin PKS system permits a
wide range of polyketides to be synthesized. As
compared to the aromatic PKS systems, a wider range of
starter units including aliphatic monomers (acetyl,
25   propionyl, butyryl, isovaleryl, etc.), aromatics
(aminohydroxybenzoyl), alicyclics (cyclohexanoyl), and
heterocyclics (thiazolyl) are found in various
macrocyclic polyketides. Recent studies have shown that
modular PKSs have relaxed specificity for their starter
30   units. Modular PKSs also exhibit considerable variety
with regard to the choice of extender units in each
condensation cycle. The degree of β-ketoreduction
following a condensation reaction has also been shown to
be altered by genetic manipulation (Donadio, S. et al.
35   Proc. Natl. Acad. Sci. USA (1993) 90:7119-7123).
Likewise, the size of the polyketide product can be
varied by designing mutants with the appropriate number

- 41 -

of modules (Kao, C. M. et al. J. Am. Chem. Soc. (1994) 116: 11612-11613). Lastly, these enzymes are particularly well-known for generating an impressive range of asymmetric centres in their products in a highly controlled manner. The polyketides and antibiotics produced by the methods of the present invention are typically single stereoisomeric forms. Although the compounds of the invention can occur as mixtures of stereoisomers, it is more practical to generate individual stereoisomers using this system. Thus, the combinatorial potential within modular PKS pathways based on any naturally occurring modular, such as the nystatin, PKS scaffold is virtually unlimited.

In general, the polyketide products of the PKS must be further modified, typically by glycosylation, in order to exhibit antibiotic activity. Methods for glycosylating the polyketides are generally known in the art; the glycosylation may be effected intracellularly by providing the appropriate glycosylation enzymes or may be effected *in vitro* using chemical synthetic means.

The macrolide antibiotics, polyketide moieties of which are synthesised by modular PKSs, may contain any of a number of different deoxysugars. The nystatin molecule contains mycosamine deoxysugar moiety. Deoxysugar biosynthesis starts typically with glucose-1-phosphate and proceeds through the action of dTDP-glucose synthase and dTDP-glucose-4,6-dehydratase. The product of the latter, typically a dTDP-4,6-keto-6-deoxyglucose, is further subjected to at least two of the following reactions - epimerisation, isomerisation, reduction, dehydration, transamination, or methylation - to give a dTDP-D-deoxysugar. The latter is then attached to the macrolactone ring via the action of a glycosyltransferase, hence providing for the glycosylation of the macrolide compound.

Glycosylation can also be effected using the non-glycosylated macrolides as starting materials, and using

- 42 -

mutants of streptomycetes or other organisms (i.e.
Myxococcus, *Pseudomonas*, *Mycobacterium* etc.) that can
provide the glycosylation activities. Alternatively,
glycosyltransferase-encoding genes from the organisms
5   mentioned above can be introduced in *S. noursei* or other
organisms containing the native or modified nystatin PKS
genes or portions thereof in order to provide the
desired glycosylation. The deoxysugar biosynthesis
genes from the nystatin gene cluster can be used for
10   complementation of corresponding activities in different
PKS producers, as well as for engineering the
biosynthetic pathways for alternative deoxysugars.

    The derivatives of nystatin PKS can be prepared by
manipulation of the relevant genes, or by introducing
15   the nystatin genes or portions thereof into another PKS.
A large number of modular PKS gene clusters have been
mapped and/or sequenced, including erythromycin,
soraphen A, rifamycin, avermectin and rapamycin, which
have been completely mapped and sequenced, and FK506 and
20   oleandomycin which have been partially sequenced, and
candicidin, pimaricin and nemadectin which have been
mapped and partially sequenced. Additional modular PKS
gene clusters are expected to be available as time
progresses. These genes can be manipulated using
25   standard techniques to delete or inactivate activity
encoding regions, insert regions of genes encoding
corresponding activities from the same or different PKS
systems, or otherwise mutate using standard procedures
for obtaining genetic alterations. Of course, portions
30   of, or all of, the desired derivative coding sequences
can be synthesized using standard solid phase synthesis
methods such as those described by Jaye et at., J. Biol.
Chem. (1984) 259:6331, and which are available
commercially from, for example, Applied Biosystems, Inc.
35   In order to obtain nucleotide sequences encoding a
variety of derivatives of the naturally occurring PKS,
and thus a variety of polyketides for construction of a

- 43 -

library, a desired number of constructs can be obtained
by "mixing and matching" enzymatic activity-encoding
portions, and mutations can be introduced into the
native host (i.e. nystatin) PKS gene cluster or portions
5   thereof.

     Mutations can be made to the native sequences using
conventional techniques. The substrates for mutation
can be an entire cluster of genes or only one or two of
them; the substrate for mutation may also be portions of
10   one or more of these genes. Techniques for mutation are
well known in the art and described in the literature,
for example in WO98/49315 and the references cited
therein. Such techniques include preparing synthetic
oligonucleotides including the mutation(s) and inserting
15   the mutated sequence into the gene encoding a PKS
subunit using restriction endonuclease digestion. (See,
e.g. Kunkel, T.A., Proc. Natl. Acad. Sci. USA (1985)
82:448; Geisselsoder et al. BioTechniques (1987) 5:786.)
Alternatively, the mutations can be effected using a
20   mismatched primer (generally 10-20 nucleotides in
length) which hybridizes to the native nucleotide
sequence, at a temperature below the melting temperature
of the mismatched duplex. The primer can be made
specific by keeping primer length and base composition
25   within relatively narrow limits and by keeping the
mutant base centrally located (Zoller and Smith, Methods
Enzymol. (1983) 100:468). Primer extension is effected
using DNA polymerase, the product cloned and clones
containing the mutated DNA, derived by segregation of
30   the primer extended strand, selected. Selection can be
accomplished using the mutant primer as a hybridization
probe. The technique is also applicable for generating
multiple point mutations. See, e.g. Dalbie-McFarland et
al. Proc. Natl. Acad. Sci. USA (1982) 79:6409. PCR
35   mutagenesis will also find use for effecting the desired
mutations.

     Random mutagenesis of selected portions of the

- 44 -

nucleotide sequences encoding enzymatic activities can
be accomplished by several different techniques known in
the art, e.g. by inserting an oligonucleotide linker
randomly into a plasmid, by irradiation with X-rays or
ultraviolet light, by incorporating incorrect
nucleotides during *in vitro* DNA synthesis, by
error-prone PCR mutagenesis, by preparing synthetic
mutants or by damaging plasmid DNA *in vitro* with
chemicals. Chemical mutagens include, for example,
sodium bisulfite, nitrous acid, nitrosoguanidine,
hydroxylamine, agents which damage or remove bases
thereby preventing normal base-pairing such as hydrazine
or formic acid, analogues of nucleotide precursors such
as 5-bromouracil, 2-aminopurine, or acridine
intercalating agents such as proflavine, acriflavine,
quinacrine, and the like. Generally, plasmid DNA or DNA
fragments are treated with chemicals, transformed into
E. coli and propagated as a pool or library of mutant
plasmids.

In addition to providing mutated forms of regions
encoding enzymatic or other functional activity, regions
encoding the desired functions or activities may be
recovered from different locations in the same nystatin
PKS, for example, using PCR techniques with appropriate
primers. By "corresponding" activity encoding regions
is meant those regions encoding the same general type of
activity -- e.g. a ketoreductase activity in one
location of a gene cluster would "correspond" to a
ketoreductase-encoding activity in another location in
the gene cluster.

If replacement of a particular target region in a
host polyketide synthase is to be made (be this host
nystatin PKS, or a different PKS into which "nystatin"
sequences are to be inserted), this replacement can be
conducted *in vitro* using suitable restriction enzymes or
can be effected *in vivo* using recombinant techniques
involving homologous sequences framing the replacement

- 45 -

gene in a donor plasmid and a receptor region in a
recipient plasmid, genome or chromosome. Such systems,
advantageously involving plasmids of differing
temperature sensitivities are described, for example, in
5   PCT application WO 96/40968.

WO 00/77181 describes methods of assembling several
DNA units in sequence into large DNA constructs which
are applicable to the recombinant polyketide synthases
within the scope of the invention.

10   The vectors used to perform the various operations
to replace the enzymatic activity in the host PKS genes
or to support mutations in these regions of the host PKS
genes may be chosen to contain control sequences
operably linked to the resulting coding sequences in a
15   manner that expression of the coding sequences may be
effected in a appropriate host. However, simple cloning
vectors may be used as well.

If the cloning vectors employed to obtain PKS genes
encoding derived PKS lack control sequences for
20   expression operably linked to the encoding nucleotide
sequences, the nucleotide sequences are inserted into
appropriate expression vectors. This need not be done
individually, but a pool of isolated encoding nucleotide
sequences can be inserted into host vectors, the
25   resulting vectors transformed or transfected into host
cells and the resulting cells plated out into individual
colonies.

Suitable control sequences include those which
function in eucaryotic and prokaryotic host cells.
30   Preferred hosts include prokaryotic hosts and fungal
systems such as yeast, but single cell cultures of, for
example, mammalian cells could also be used. There is
no particular advantage, however, in using such systems.
Particularly preferred are yeast and prokaryotic hosts
35   which use control sequences compatible with *Streptomyces*
spp. Suitable control sequences for single cell
cultures of various types of organisms are well known in

- 46 -

the art. Systems for expression in yeast, including
control sequences which effect secretion are widely
available and are routinely used. Control elements
include promoters, optionally containing operator

5   sequences, and other elements depending on the nature of
the host, such as ribosome binding sites. Particularly
useful promoters for prokaryotic hosts include those
from PKS gene clusters which result in the production of
polyketides as secondary metabolites, including those

10   from aromatic (Type II) PKS gene clusters. Examples are
act promoters, tcm promoters, spiramycin promoters, and
the like. However, other bacterial promoters, such as
those derived from sugar metabolizing enzymes, such as
galactose, lactose (tac) and maltose, are also useful.

15   Additional examples include promoters derived from genes
encoding biosynthetic enzymes such as tryptophan
synthase (trp), the β-lactamase (bla), bacteriophage
lambda FL, T5 and T7. In addition, synthetic promoters,
such as the tac promoter (U.S. Patent No. 4,551,433),

20   can be used.

       Other regulatory sequences may also be desirable
which allow for regulation of expression of the PKS
replacement sequences relative to the growth of the host
cell. Regulatory sequences are known to those of skill

25   in the art, and examples include those which cause the
expression of a gene to be turned on or off in response
to a chemical or physical stimulus, including the
presence of a regulatory compound. Other types of
regulatory elements may also be present in the vector,

30   for example, enhancer sequences.

       Selectable markers can also be included in the
recombinant expression vectors. A variety of markers
are known which are useful in selecting for transformed
cell lines and generally comprise a gene whose

35   expression confers a selectable phenotype on transformed
cells when the cells are grown in an appropriate
selective medium. Such markers include, for example,

- 47 -

genes which confer antibiotic resistance or sensitivity
to the plasmid. Alternatively, several polyketides are
naturally coloured and this characteristic provides a
built-in marker for screening cells successfully
5   transformed by the present constructs.

The various PKS nucleotide sequences, or a mixture
of such sequences, can be cloned into one or more
recombinant vectors as individual cassettes, with
separate control elements, or under the control of, e.g.
10   a single promoter. The PKS subunits or mixture of
components can include flanking restriction sites to
allow for the easy deletion and insertion of other PKS
subunits or mixture components so that hybrid PKSs can
be generated. The design of such unique restriction
15   sites is known to those of skill in the art and can be
accomplished using the techniques described above, such
as site-directed mutagenesis and PCR.

As described above, particularly useful control
sequences are those which themselves, or using suitable
20   regulatory systems, activate expression during
transition from growth to stationary phase in the
vegetative mycelium. The system contained in plasmid
RM5, i.e. the *actI/actIII* promoter pair and the *actII-
ORF4*, an activator gene, is particularly preferred
25   (McDaniel et al., Science, v. 262, p 1546-1550, 1993).
Particularly preferred hosts are those which lack their
own means for producing polyketides so that a cleaner
result is obtained. Illustrative host cells of this type
include the modified *S. coeticotor* CH999 culture
30   described in PCT application WO 96/40968 and similar
strains of *S. lividans*.

The expression vectors containing nucleotide
sequences encoding a modified PKS system or a variety of
PKS systems for the production of different polyketides
35   may then be transformed into the appropriate host cells,
e.g. to construct a library. In one straightforward
approach, a mixture of such vectors is transformed into

- 48 -

the selected host cells and the resulting cells plated into individual colonies and selected for successful transformants. Each individual colony will then represent a colony with the ability to produce a particular PKS synthase and ultimately a particular polyketide. Typically, there will be duplications in some of the colonies; the subset of the transformed colonies that contains a different PKS in each member colony can be considered the library. Alternatively, the expression vectors can be used individually to transform hosts, which transformed hosts are then assembled into a library. A variety of strategies might be devised to obtain a multiplicity of colonies each containing a PKS gene cluster derived from the nystatin host gene cluster so that each colony in the library produces a different PKS and ultimately a different polyketide. The number of different polyketides that are produced by the library is typically at least three (e.g. 2 mutations in the PKS genes which may appear separately or in combination), more typically at least ten, and preferably at least 20, more preferably at least 50, reflecting similar numbers of different altered PKS gene clusters and PKS gene products. The number of members in the library is arbitrarily chosen; however, the degrees of freedom outlined above with respect to the variation of starter, extender units, stereochemistry, oxidation state, and chain length is quite large.

Methods for introducing the recombinant vectors of the present invention into suitable hosts are known to those of skill in the art and typically include the use of $CaCl_2$ or other agents, such as divalent cations, lipofection, conjugation, protoplast transformation and electroporation.

A wide variety of hosts can be used, even though some hosts natively do not contain the appropriate post-translational mechanisms to activate the acyl carrier

- 49 -

proteins of the synthases. These hosts can be
complemented with the appropriate recombinant enzymes,
for example NysF, to effect these modifications.

5   The polyketide producing colonies can be identified
and isolated using known techniques and the produced
polyketides further characterized. The polyketides
produced by these colonies can be used collectively in a
panel to represent a library or may be assessed
individually for activity.

10   The libraries can thus be considered at four
levels: (1) a multiplicity of colonies each with a
different PKS encoding sequence comprising a different
PKS cluster but all derived from the nystatin PKS
cluster; (2) colonies which contain the proteins that

15   are members of the PKS produced by the coding sequences;
(3) the polyketides produced; and (4) antibiotics
derived from the polyketides.

       Colonies in the library can be induced to produce
the relevant synthases and thus to produce the relevant

20   polyketides to obtain a library of candidate
polyketides. The polyketides produced can be screened
for antimicrobial, antitumour, antihelmintic or
immunosuppressive activities, as well as for binding to
desired targets, such as receptors, signalling proteins,

25   and the like. The supernatants or culture pellets per
se can be used for screening, or partial or complete
purification of the polyketides can first be effected.
Typically, such screening methods involve detecting the
binding of each member of the library to receptor or

30   other target ligand. Binding can be detected either
directly or through a competition assay. Means to
screen such libraries for binding are well known in the
art.

       Alternatively, individual polyketide members of the
35   library can be tested against a desired target. In this
event, screens wherein the biological response of the
target is measured can more readily be included.

- 50 -

A large number of novel polyketides may be prepared according to the method of the invention, as illustrated by the representative in the Examples below. These novel polyketides may be useful intermediates in
5   formation of compounds with antibiotic activity through glycosylation or reactions as described above. As indicated above, the individual polyketides may be reacted with suitable sugar derivatives to obtain compounds of antibiotic activity. Antibiotic activity
10   can be verified using typical screening assays such as those set forth in Lehrer, R. et al., J. Immunol. Meth. (1991) 137:167-173.

Thus, in a further aspect, the invention provides a method of preparing a nucleic acid molecule comprising a
15   nucleotide sequence encoding a modified nystatin PKS, wherein said modified nystatin PKS is derived from a nystatin PKS as defined herein (i.e. a naturally occurring nystatin PKS) encoded by a nucleotide sequence as defined herein) containing first regions which encode
20   enzymatic or other functional activities and second regions which encode scaffolding amino acid sequences, said method comprising

(a)   modifying at least one said first region; or

(b)   incorporating at least one said first region
25   into a scaffolding-encoding second region from a different PKS-encoding nucleotide sequence.

As discussed above in relation to the nucleic acid molecules of the invention, the first region may be any part (e.g. encoding a domain or a module or a part
30   thereof) of a nucleotide sequence of the invention (i.e. SEQ ID NO 1 or 2 or 35).

Also provided are (i) a method of preparing a modified nystatin PKS as defined above, said method comprising expressing a nucleic acid molecule prepared
35   as defined above within a host cell (i.e. culturing or growing a host cell containing such a nucleic acid molecule) under conditions whereby the modified nystatin

- 51 -

PKS is expressed), and (ii) a modified nystatin PKS so produced or so obtainable.

New polyketides produced by such a modified PKS are also within the scope of the invention, as are new
5   antibiotics which are generally the glycosylated forms of these polyketides although some have activity without glycosylation which may be due to different post-translation modifications such as hydroxylation or oxidation.

10   The invention will now be described in more detail in the following non-limiting Examples with reference to the drawings in which:

    Figure 1 shows the structure of the polyene antifungal antibiotic nystatin A1;

15   Figure 2 presents physical and functional maps of the *E.coli* - *Streptomyces* shuttle vector pSOK101, pSOK201 and pSOK804 used in Examples 1, 2 and 3;

    Figure 3 is a schematic representation showing two regions of the *S.noursei* ATCC11455 genome encoding the
20   nystatin biosynthesis gene (corresponding to SEQ ID NOs 1 and 2). Overlapping recombinant phages containing the presented DNA sequences are shown over the regions drawings (see Example 1);

    Figure 4 is a schematic representation showing the
25   functional organisation of the nystatin PKS NysA (SEQ ID NO 5), NysB (SEQ ID NO 6), NysC (SEQ ID NO 7) and NysI (SEQ ID NO 20) and their roles in nystatin biosynthesis;

    Figure 5 is a schematic representation showing genetic manipulations of the module 5 in NysC PKS
30   leading to production of new polyene compounds by recombinant *S.noursei* strains (see Example 2);

    Figure 6 shows the UV spectra (in DMSO) for nystatin and new polyene compounds S44 and S48 obtained from recombinant *S. noursei* strains with genetically
35   altered NysC PKS (see Example 2);

    Figure 7 is a schematic representation showing the region of the *S. noursei* ATC11455 genome encoding the

- 52 -

nystatin biosynthetic gene cluster (corresponding to SEQ
ID No. 35). Gene organisation within the gene cluster
is shown. The inserts from the overlapping recombinant
phages encompassing the cloned region are shown above
the physical/genetic map. The *nys* genes are designated
with capital letters in *italic*, other ORFs are numbered;

Figure 8 is a representation showing a proposed
model for nystatin biosynthesis in *S. noursei*; and

Figure 9 is a schematic representation showing the
functional organisation of the nystatin PKS NysA (SEQ ID
No. 5), NysB (SEQ ID No. 6), NysC (SEQ ID No. 7), NysI
(SEQ ID No. 37), NysJ (SEQ ID No. 38) and NysK (SEQ ID
No. 39) proteins. KS^s - ketosynthase with the Cys to Ser
substitution in active site; KS - ketosynthase; AT -
acetate-specific acyltransferase; mAT - propionate-
specific acetyltransferase; DH - dehydratase; DHi -
inactive dehydratase; ER - enoyl reductase; KR -
detoreductase; KRi - inactive ketoreductase; ACP - acyl
carrier protein.

Figure 10 shows compounds that can be theoretically
produced from the following manipulations within the
nystatin gene cluster
-insertion of ER domain into module 3 (1)
-insertion of ER domain into module 4 (2);
-simultaneous inactivation of the ER domain in module 5
and insertion of the ER domain into module 3 (3);
-simultaneous inactivation of the ER domain in module 5
and insertion of the ER domain into module 4 (4);
-simultaneous inactivation of the ER domain in module 5
and insertion of the ER domain into module 7 (5);
-simultaneous inactivation of the ER domain in module 5
and insertion of the ER domain into module 8 (6);
-simultaneous inactivation of the ER domain in module 5
and insertion of the ER domain into module 9 (7);
-simultaneous inactivation of the ER domain in module 5
and insertion of the ER domains into modules 8 and 9
(8);

- 53 -

-simultaneous inactivation of the ER domain in module 5
and insertion of the ER domains into modules 7 and 8
(9).

<u>Figure 11</u> shows compounds that can be theoretically
5   produced from the following manipulations within the
nystatin gene cluster:
-replacement of methylmalonyl-specific acetyltransferase
(AT) domain in module 11 of the nystatin PKS with
malonyl-specific AT domain (10);
10   -replacement of malonyl-specific AT domain in module 12
with methylmalonyl-specific AT domain with simultaneous
replacement of methylmalonyl-specific AT domain in
module 11 with malonyl-specific AT domain (11);
-replacement of malonyl-specific AT domain in module 10
15   with methylmalonyl-specific AT domain with simultaneous
replacement of methylmalonyl-specific AT domain in
module 11 with malonyl-specific AT domain (12);
-inactivation of P450 monooxygenase-encoding genes nysL
or nysN (whichever is found to be responsible for
20   oxygenation of the methyl group at C-16 on the nystatin
molecule) (13).

<u>Figure 12</u> shows compounds that can be theoretically
produced from the following manipulations within the
nystatin gene cluster:
25   -inactivation of dehydratase (DH) domain in module 3 of
the nystatin PKS (14);
-inactivation of DH domain in module 4 (15);
-inactivation of DH domain in module 3 with simultaneous
inactivation of ER domain in module 5 (16);
30   -inactivation of DH domain in module 4 with simultaneous
inactivation of ER domain in module 5 (17) ;
-inactivation of DH domain in module 7 with simultaneous
inactivation of ER domain in module 5 (18);
-inactivation of DH domain in module 8 with simultaneous
35   inactivation of ER domain in module 5 (19);
-inactivation of DH domain in module 9 with simultaneous
inactivation of ER domain in module 5 (20).

- 54 -

Example 1 - Cloning of the nystatin biosynthesis gene cluster

Bacterial strains, plasmids and growth conditions

Bacterial strains and plasmids used in this study
are listed in Table 3. New strains and plasmids
developed in the course of this study are described
herein and shown in Figure 3. *S.noursei* ATCC 11455 and
its mutants were grown on solid ISP2 medium (Difco), and
in liquid TSB medium (Oxoid). Intergeneric conjugation
from *E.coli* ET12567 (pUZ8002) into *Streptomyces* strains
was done as described in Flett *et al*, FEMS Microbial
Lett., v. 155: 223-229, (1997), except for the "heat
shock" time, which was reduced to 5 minutes. Apramycin
(Fluka) at a concentration 50 mg/ml was used to select
for the *S.noursei* transconjugants on solid medium. For
inoculation of the *S.noursei* ATCC11455 transconjugants
prior to total DNA isolation, liquid medium TSB
supplemented with 20 μg/ml apramycin was used. *E.coli*
strains were grown and transformed as described in
Sambrook *et al*, Cold Spring Harbor Laboratory (1989).
E.coli ET12567 (pUZ8002) was maintained on the media
containing 20μg/ml chloramphenicol and 50 μg/ml
kanamycin.

Table 3 - Bacterial strains, plasmids and phages used in this study

Strain, plasmid or phage	Properties	Source or reference
E.coli DH5α	general cloning host	Sambrook *et al*, 1989, supra
E.coli XL1-Blue MRA (P2)	host for the gene library	Stratagene
E.coli ET12567	strain for intergeneric conjugation	MacNeil *et al*, Gene, v. 111: 61-68, 1992
S.noursei ATCC 11455	wild type, nystatin producer	ATCC

- 55 -

pGEM3Zf(-)	ColE1 replicon, Ap^R, 3.2kb	Promega
pGEM72f(-)	ColE1 replicon, Ap^R, 3.0 kb	Promega
5 pGEM11Zf(-)	ColE1 replicon, Ap^R, 3.2kb	Promega
pUZ8002	RK2 derivative, Km^R, Tc^R	D.H.Figurski
pWHM3	ColE1+pIJ101 replicons, Thio^R, 7.2kb	Vara et al, J.Bacteriol v. 171: 5872-5881, 1989
pSET152	ColE1 replicon+øC31 int, oriT,Am^R, 5.5kb	Bierman et al, Gene, v. 116: 43-49, 1992
pSOK101	pWHM3 derivative in which the 3.1kb BamHI/SphI fragment was replaced with the 3.0 kb BamHI/SphI fragment from pSET152 containing ColEI, oriT and Am^R, 7.1kb	This work (see Figure 2)
10 pGM11	pSG5 replicon, Neo^R, 5.3kb	Wohlleben & Muth, In "Plasmids, a practical approach", Ed. Hardy, IRL Press, p 147-175, 1993
pSOK201	pGM11 derivative in which the 1.2 kb EcoRI/HindIII fragment was replaced with the 3.0 kb EcoRI/HindIII fragment from pSOK101 containing ColEI, oriT and Am^R, 7.1 kb	This work (see Figure 2)

- 56 -

DASHII	bacteriophage λ vector	Stratagene
pSOK804	ColEI replicon, Am^R, *OriT*, int$_{WVB}$, AttP$_{WVB}$	This work (see Figure 2) and Van Mellaert et al., Microbiol., v. 144:3351-3358, 1998
pGEM7ermELi	pGEM7ZF plasmid containing PermE* promoter	C.R. Hutchinson

Am-apramycin, Ap-ampicilli, Neo-neomycin, Thio-thiostrepton, Km-kanamycin, Tc-tetracycline

Analysis of the secondary metabolite production

    Fermentations were performed in Applicon 3-1 fermenters containing initially 1.3 1 SAO-23 medium containing (gl^{-1}):glucose·H_2O, 90; NH_4NO_3, 2.5; corn flour, 3.0; $MgSO_4$·$7H_2O$, 0.4; KH_2PO_4,02; $CaCO_3$, 7; trace element solution, 3ml. Trace element solution (gl^{-1}):$FeSO_4$.$7H_2O$, 5.0; $CuSO_4$.$5H_2O$, 0.39; $ZnSO_4$$7H_2O$, 0.44; $MnSO_4$$H_2O$, 0.15; Na_2MoO_4.$2H_2O$, 0.01; $CoCl_2$.$6H_2O$, 0.02; HCl, 50. The fermentations were performed at 28°C with pH controlled at 6.5-7.0 by HCl (2M) and NaOH (2M). The dissolved oxygen was controlled at >40% of saturation by the agitation (300-900 rpm) and aeration (0.25 vvm). Inocula for the fermentations (3 vol-%) with SAO-23 medium were grown in TSB-medium (TSB, Oxoid CM129, 37 gl^{-1}) at 28°C in shake flasks (500-ml baffled Erlenmeyer flasks with 100 ml medium; 200 rpm). Each shake flask was inoculated with 0.2 ml spore suspension and incubated for 18-20 hours. Nystatin production was assayed by HPLC of the dimethylformamide extracts of the cultures after fermentations (Raatikainen, J. Chromatogr., v. 588, 356-360, (1991)).

- 57 -

DNA Manipulation

Plasmid, phage and total DNA preparations, endonuclease digestions, ligations and fractionation were performed as described previously (Sambrook *et al.*, 1989, supra; Hopwood *et al*, 1985, supra). DNA fragments were isolated from agarose gels using the QIAGEN Kit (QIAGEN GmbH), labelled with the use of the digoxygenin kit from Boehringer Mannheim and used for Southern blot analysis according to the manufacturer's manual. DNA sequencing was performed at QIAGEN GmbH (Germany), and the data were analysed with the GCG software (Devereux *et al*, Nucleic Acids Res. v. 12: 387-395, 1984).

PCR-assisted amplification and cloning of a PKS-encoding DNA fragment from the S.noursei ATCC11455 genome

In order to obtain the DNA encoding the nystatin biosynthesis genes, the *S.noursei* ATCC1455 gene library was probed with labelled PKS-encoding DNA. To obtain the DNA probe, degenerate oligonucleotide primers were designed, which correspond to conserved amino acid regions within β-ketoacyl synthase (KS) and acyl carrier protein (ACP) domains of known modular PKSs. The degenerate primers used for amplification corresponded to the conserved amino acid motifs in ACP and KS domains in known PKS, and were designed according to the codon usage table for *Streptomyces* (Wright & Bibb, Gene, v. 113: 55-65, 1992). The ACP nucleotide primer (sense) corresponded to the motif Glu(Asp)Leu Gly Phe(Leu, Val) Asp Ser Leu (SEQ ID NO:21) and had the sequence 5' - GAG/C CTG/C GGC/G T/CTG/C GAC TCC/G CTG/C-3 (SEQ ID NO: 22). The KS nucleotide primer (antisense) corresponded to the motif Val Asp Thr Ala Cys Ser Ser (SEQ ID NO: 23) and had the sequence 5' - G/CGA G/CGA G/ACA/ G/CGC C/GGT GTC G/CAC-3' (SEQ ID NO: 24). Total DNA isolated from the *S.noursei* ATCC11455 was used as a template for polymerase chain reaction (PCR)-assisted amplification of the DNA fragment from the genome of this organism

- 58 -

with the use of KS and ACP oligonucleotide primers.
From the relative position of the motifs on the modular
PKSs it was assumed that resulting PCR product would be
approx 0.7 kb in size. The 50µl PCR mixture contained
5   0.1 µg of S.noursei ATCC11455 total DNA, 25 pm of each
ACP and KS oligonucleotide primers dNTPs (final
concentration 350 µm), 1xPCR buffer from Expand High
Fidelity PCR System (Boehringer Mannheim), and 1.5 U of
the DNA polymerase mixture from the same system. The
10   PCR was performed on the Perkin Elmer GeneAmp PCR System
2400 with the following program: 1 cycle of denaturation
at 96°C (4 min), 35 cycles of denaturation/annealing/
synthesis at 94°C (45 sec) and 70°C (5 min) and 1 cycle
of final annealing/extension at 72°C (7 min). The 0.7
15   kb DNA fragment obtained with this procedure was cloned
in pUC18 vector in E.coli DH5α with the use of Sure
Clone Ligation Kit (Pharmacia). One of the resulting
recombinant plasmids, pPKS72 of 3.5kb, was subjected to
DNA sequence analysis.

20   Subsequent cloning in Escherichia coli vector pUC18
and DNA sequencing of the resulting 0.7 kb PCR product,
followed by conceptual translation and database search,
confirmed that it encodes part of PKS type I. This DNA
fragment was used as a probe for screening a S.noursei
25   ATCC11455 gene library constructed in the phage vector
DASHIII (see below) and one recombinant phage,
designated lambda DASHII-N1, which hybridized to the
probe, was isolated. As described further below,
DASHII-N1 was used to generate further probes.

30

Construction and screening of the S.noursei ATCC1455
gene library
        The S.noursei ATCC11455 gene library was
constructed in phage lambda vector DASHII (Stratagene)
35   according to manufacturer's instructions. Total DNA
from S.noursei ATCC11455 was isolated as described in
Hopwood et al (1985) supra, partially digested with Sau3

- 59 -

AI restriction enzyme and fractionated on the sucrose
gradient as described in Sambrook *et al* (1989), supra.
The fractions of *S.noursei* ATCC1455 DNA containing
fragments of 13-17 kb in size were used for ligation
5   with the DASHII vector arms digested with *Bam*HI
restriction enzyme. *E.coli* XL1-Blue MRA (P2)
(Stratagene) was used as a host for a gene library
construction and propagation.
        DNA fragments to be used as probes for screening
10   the gene library were purified from agarose gels using
QIAGEN Kit (QIAGEN GmbH, Germany), and labelled by the
use of the digoxygenin (DIG) kit from Boehringer
Mannheim (Germany), according to the manufacturer's
instructions. Probes used for the library screening and
15   relevant recombinant phages discovered:

Probe	DASHII recombinant phages found using this probe
PKS72	N1, N14
E12.1: from N1	N41, N42, N44, N45, N48
E4.7.2: from N1	N58
L42E9.1: from N42	N64, N76
B1.0_58: from N58	N69

Description of the probes:

25   1.   PKS72 probe. The 0.7kb DNA fragment isolated from
the pPKS72 plasmid (see above) with restriction enzymes
*Eco*RI and *Hin*dIII was used as a PKS72 probe.
      2.   E12.1 probe. The 2.6 kb *Bam*HI fragment from the
insert of recombinant phage N1, representing its left
30   flanking region, was subcloned into pGEM3Zf(-),
resulting in plasmid pGEM(B2.6)-1. This plasmid was
digested with *Eco*RI/*Ava*I, and the 0.55 kb fragment
corresponding to the left end of the N1 DNA insert was
purified and used as E12.1 probe.
35   3.   E4.7.2 probe. The 4.7 kb *Eco*RI fragment from the

- 60 -

recombinant phage N1, representing its right flanking region was subcloned into pGEM3Zf(-), resulting in plasmid pGEM (E4.7)-1. This plasmid was digested with EcoRI/HindIII, and the 1.5kb DNA fragment corresponding

5   to the right end of the N1 DNA insert was purified and named E4.7.2.

4.   L42E9.1 probe. The 9.0 kb EcoRI fragment representing the left flanking region of the DNA insert in the recombinant phage N42 was subcloned into

10 pGEM3Zf(-), resulting in plasmid pH42E9.1. This plasmid was digested with EcoRI/BamHI, and the 0.6 kb fragment corresponding to the left flank of the N42 DNA insert was purified and used as L42E9.1 probe.

5.   L58B1.0 probe. The 3.0 kb EcoRI fragment

15 representing the right flanking region of the recombinant phage N58 DNA insert was subcloned into pGEM3Zf(-), resulting in plasmid pGEM(E3.0)-58. From the latter plasmid, the 1.0 kb BamHI fragment which is located on the right end of the N58 DNA insert was

20 purified and used as L58B1.0 probe.

Gene disruption experiment with the nystatin biosynthesis gene cluster

A 4.2 kb BamHI DNA fragment isolated from the

25 DASHII-N1 recombinant phage was first cloned into the pGEM3Zf(-) vector in E.coli, resulting in the plasmid pGEM 4.2-1. DNA sequences on both ends of the cloned fragment were determined and, after database search, were found to encode PKS type I. The S.noursei DNA

30 fragment cloned in pGEM4.2-1 was excised from this plasmid with restriction enzymes EcoRI and HindIII, and ligated with the 3.0 kb EcoRI/HindIII fragment from the vector pSOK201 (see Table 3 and Figure 3), and transformed into E.coli. Plasmid DNA designated

35 pKO(4.2)-1 was recovered from transformants, and then was transferred into S.noursei ATCC1455 by conjugation as described in Example 1 under "Bacterial strains,

plasmids and growth conditions".

    Since pKO(4.2)-1 is not capable of replicating in
S.noursei ATCC11455, it was assumed that it will
function as a suicide vector integrating into the genome

5   of S.noursei via homologous recombination. As a result
of such recombination, a PKS gene in the S.noursei
ATCC11455 genome, for which 4.2 kb BamHI fragment cloned
in pKO(4.2)-1 was presumed to be internal, would have
been inactivated by disruption of its coding region.

10   Integration of pKO(4.2)-1 into the genome of the three
S.noursei transconjugants was confirmed by Southern blot
analysis with the use of labelled 4.2 kb BamHI fragment
from pGEM4.2-1 as a probe. One of the S. noursei
disruption mutants carrying pKO(4.2)-1 integrated into

15   its genome was tested for nystatin production in
parallel with the parental strain ATCC11455 (see above
for methods under "analysis of secondary metabolite
production". While the latter was shown to produce
nystatin at the expected level, no nystatin production

20   was detected in the pKO(4.2)-1 disruption mutant, thus
confirming the requirement of the identified PKS for
nystatin biosynthesis.

Cloning of the nystatin biosynthesis gene cluster

25   In order to clone the entire gene cluster for the
nystatin biosynthesis, the DNA fragments derived from
the ends of the S. noursei DNA insert in the recombinant
phage DASHII-N1 (N1), and subsequently found overlapping
recombinant phages, were used as probes for screening

30   the gene library (see above for probes). This screen
resulted in isolation of the recombinant phages N14,
N41, N42, N44, N45, N48, N58, N64, N69, and N76
comprising two regions (SEQ ID NOs 1 and 2 respectively)
of the S. noursei ATCC11455 genome (approx. 98 kb

35   total), as depicted in Fig. 4. A gene disruption
experiment with the 4.3 kb EcoRI/BamHI DNA fragment
derived from the recombinant phage N64 (performed

- 62 -

essentially the same way as described above), confirmed
that the second region (SEQ ID NO. 2) also encodes
nystatin biosynthesis genes.

*DNA sequence analysis of the nystatin biosynthesis gene
cluster*

The complete DNA inserts from recombinant phages
mentioned above were subcloned either as *XbaI* or *EcoRI*
fragments into pGEM3Zf(-) vector in an *E. coli* host, and
nucleotide sequences were determined on both DNA strands
of these fragments. Computer-assisted analysis of the
DNA sequences comprising the two regions of the nystatin
biosynthesis gene cluster (SEQ ID NOs 1 and 2
respectively) resulted in identification of the genes
shown on Fig. 4 and listed in Table 4.

Table 4. Genes identified within the nystatin
biosynthesis gene cluster of *S.noursei*

Designation	Product	Putative function
nysR1	transcriptional activator	regulation of nystatin production
nysR2	transcriptional activator	regulation of nystatin production
nysR3	transcriptional activator	regulation of nystatin production
nysR4	transcriptional activator	regulation
nysR5	transcriptional repressor	regulation
ORF1	peptidase	peptide metabolism
ORF2	transcriptional activator	regulation
nysA	PKS type I	nystatin polyketide backbone synthesis (loading module)
nysB	PKS type I	nystatin polyketide backbone synthesis (modules 1&2)

- 63 -

nysC	PKS type I	nystatin polyketide backbone synthesis (modules 3-8)
nysI (incomplete)	PKS type I	nystatin polyketide backbone synthesis (modules 9-13)
nysE	thioesterase	release of final product from PKS
nysD1	glycosyltransferase	attachment of mycosamine to the polyketide backbone
nysD2	aminotransferase	mycosamine biosynthesis
nysD3	GDP-mannose-4,6-dehydratase	mycosamine biosynthesis
nysH	ABC transporter	efflux of nystatin
nysG	ABC transporter	efflux of nystatin
nysF	4'-phosphopantetheine transferase	post-translational modification of PKS

Three complete (*nysA, nysB and nysC*)(in SEQ ID NO:1), and one incomplete (*nysI*) genes (in SEQ ID NO:2) encoding the PKSs type I were identified. The amino acid (aa) sequences of the products encoded by these four genes were analysed by comparison to the aa sequences of known PKSs type I (see also Table 2 above for molecule features). Since all four proteins displayed high degree of homology towards rifamycin and rapamycin PKSs (Aparicio et al., Gene, v. 169: 9-16, 1996; Tang et al., Gene, v. 216: 255-265, 1998), presumptive functional analysis of nystatin PKSs was based on comparison to the formers. The NysA protein of 1366 aa encodes one module of PKS composed of KS^s, AT, DH, KR, and ACP domains. The lack of a conserved cysteine residue in KS^s domain suggests that this module cannot perform condensation reaction, and thus most probably represents a loading module providing the acetate starter unit for initiation of nystatin

- 64 -

polyketide backbone biosynthesis (Bisang et al., Nature, v. 401: 502-505, 1999). Analysis of the 3192 aa sequence of NysB revealed that it contains two modules with KS, AT, KR, ACP, and apparently inactive DH

5 domains. The AT domains identified in both NysB modules display features characteristic for the methylmalonylCoA-specific AT domains (Haydock et al., *FEBS Lett.*, v. 374: 246-248, 1995). This feature of the NysB protein suggests that it comprises 1st and 2nd

10 modules involved in the nystatin polyketide backbone biosynthetic pathway, as the only two proximal methylmalonyl CoAs incorporated in nystatin molecule are the first two extender units. NysC protein of 11096 aa, the largest, to our knowledge, bacterial polypeptide

15 discovered so far, is composed of 6 modules apparently responsible for the condensation steps 3 to 8 in nystatin polyketide chain biosynthesis (incorporation of C21 - C32). Module 5 of the NysC protein contains an ER domain, which is accountable for the reduction of the

20 double bond between C29 and C28 (see Fig. 1). Besides module 5, all other modules in NysC are similar in that they all contain KS, AT, DH, KR and ACP domains. It was noticed that KR domains in modules 4 and 5 are 100% identical on the aa sequence level, and 99.9% identical

25 at the level of DNA sequences encoding these domains. Thus, KR domains in modules 4 and 5 most probably represent an example of a relatively recent duplication in the process of evolution. NysI C-terminally truncated protein of 7066 aa, for which aa sequence

30 information at the C-terminus is still missing, is composed of at least 4 complete modules. All these modules contain KS, AT, DH, KR and ACP domains, but the DH domains in three modules are apparently inactive, suggesting that the known part of NysI PKS is

35 responsible for the elongation steps 9 through 12 (incorporation of C13 - C20). This assumption is further supported by the fact that the AT domain in

- 65 -

module 11 (based on sequence similarity) is specific for
methylmalonylCoA extender, while all other AT domains in
NysA, NysC and NysI are malonyl CoA-specific. Thus, the
AT domain in module 11 is presumably responsible for the
5   occurrence of the methyl group at C16 on the nystatin
polyketide backbone, which is later oxidized to give a
C16-coupled carboxyl (see Fig. 1).

Downstream of the nysC gene, a coding sequence nysE
for thioesterase, presumably responsible for the release
10   of mature polyketide chain from the nystatin PKSs, was
found. The nysR1, nysR2 and nysR3 genes, were found
downstream of nysE. The products of these genes are
homologous to the presumed transcriptional regulators.
NysR1 (966 aa), NysR2 (953 aa), and NysR3 (927 aa)
15   proteins were all found to contain putative
helix-turn-helix (HTH) DNA binding motifs of LuxR type
at their C-termini. Beside that, NysR1 contained a
distinct ATP/GTP binding motif, and NysR3 contained a
"leucine zipper" (putative DNA binding) motif at their
20   N-termini. The gene encoding NysR1 was found to contain
a rare TTA codon (for Leu11) close to the beginning of
the gene, suggesting that NysR1 expression might be
regulated in S. noursei at the level of translation by a
bldA-like gene (White & Bibb, J. Bacteriol, v. 179: 627-
25   633, 1997).

In order to confirm the involvement of nysR1 gene
in the nystatin biosynthesis a gene disruption
experiment was performed with the use of a 1379 bp ApaI
DNA fragment internal for the nysR1 coding sequence.
30   The 1379 bp ApaI DNA fragment internal for the nysR1
coding sequence and representing nt 51531-52910 of SEQ
ID NO 1 was cloned into the ApaI site of the pGEM11Zf(-)
vector, giving the recombinant plasmid pNRD1. The 1430
bp EcoRI/HindIII DNA fragment isolated from pNRD1 was
35   ligated with the 3.0 kb EcoRI/HindIII fragment of
pSOK201 resulting in the pNRD2 vector. The latter was
subsequently used for nysR1 gene disruption in S.

- 66 -

noursei ATCC11455 via conjugation from *E. coli* ET12567
(pUZ8002).

Analysis of the nystatin production by the *nysR1*
disruption mutant revealed that it is not capable of
producing nystatin. It shall be noted, however, that
the phenotype observed for the *nysR1* disruption mutant,
might reflect the polar effect of mutation on the *nysR2*
and *nysR3* genes, which can be co-transcribed with *nysR1*.

Downstream of *nysR3*, the genes *nysR4*, *nysR5*, ORF1,
and ORF2 were found. The *nysR4* gene product of 210 aa
(this has subsequently found to be 266 aa expression
from an upstream start codon at nucleotide 120628 in SEQ
ID No. 35) shows similarity to transcriptional
activators of response regulator type, and contains
centrally located ATP/GTP binding and C-terminally
located LuxR-type HTH DNA binding motifs. NysR5 protein
of 253 aa displays similarity to the transcriptional
repressors, and contains a putative DeoR-type HTH DNA
binding motif at its N terminus. It seems likely that
nysR4 and *nysR5* gene products are involved in regulation
of nystatin biosynthesis based on their location
proximal to the nystatin biosynthesis genes. ORF2,
located downstream of *nysR5*, and transcribed in the
opposite direction, encodes a 354 aa peptide showing
similarity to transcriptional activators and having
centrally located putative HTH DNA binding motif of AsnC
type. Whether this gene is involved in regulation of
nystatin biosynthesis is not apparent, as ORF1, located
immediately upstream of ORF2, encodes a putative
peptidase. It seems likely that ORF2 is rather involved
in regulation of ORF1, but to confirm this,
experimentation on ORF2 inactivation is required. The
fact that the gene encoding a peptidase, for which no
role in nystatin biosynthesis could be assigned, was
found on the right flank of the sequenced region,
suggests that the right border of the nystatin
biosynthesis gene cluster had been identified.

- 67 -

On the left flank of the sequenced region
encompassing the genes described above, two genes
located upstream of *nysA*, and transcribed in the
direction opposite to *nysA* were identified (Fig. 4).
5   The *nysD1* gene product of 506 aa displays considerable
homology to the UDP-glucuronosyltransferases from
mammals. This enzyme belongs to the
UDP-glycosyltransferase family, and takes part in the
process of elimination of potentially toxic xenobiotics
10   by the way of their glycosylation. It seems likely that
NysD1 represents a glycosyltransferase responsible for
the attachment of the deoxysugar moiety (mycosamine) to
the nystatin polyketide backbone. The product of *nysD2*
shows a high degree of homology to the perosamine
15   synthetases and transaminases responsible for the
attachment of amino groups to the deoxysugars in the
process of their biosynthesis. Thus, NysD2 presumably
represents an aminotransferase involved in mycosamine
biosynthesis.
20   Beside *nysI* encoding PKS type I, four other genes
were identified within the second sequenced region (SEQ
ID NO. 2) of the nystatin biosynthesis gene cluster of
S. noursei ATCC11455 (Fig. 4). The 344 aa protein
encoded by the *nysD3* gene is highly homologous to the
25   GDP-mannose-4,6-dehydratases required for deoxysugar
formation. It is thus likely that the NysD3 protein is
involved in biosynthesis of the nystatin mycosamine
moiety in *S. noursei* ATCC11455. *nysH* and *nysG* gene
products of 584 aa and 605 aa, respectively, display a
30   high degree of similarity to transporters of the ABC
family. Both NysH and NysG polypeptides contain a
distinct ABC transporter signature at their C-termini,
centrally located ATP/GTP binding motifs, and
N-terminally located transmembrane regions. These two
35   proteins most probably are responsible for the
ATP-dependent active efflux of nystatin from the
producing organism, thus eliminating the danger of

- 68 -

membrane clogging by the hydrophobic nystatin molecules.
The product of the *nysF* gene, a 245 aa polypeptide,
displays homology to the 4'-phosphopantetheine
transferases. The latter enzyme is responsible for the
post-translational modification of the ACP domains of
the PKSs, and is required for its full functionality
(Cox et al., *FEBS Lett.*, v. 405: 267-272, 1997; Kealey
et al., Proc. Natl. Acad. Sci. USA, v. 95: 505-509,
1998). It seems likely, therefore, that the NysF
protein functions in post-translational modification of
nystatin PKSs, and is required for nystatin
biosynthesis.

Example 2 - Genetic manipulation of the nystatin PKS genes leading to production of novel polyene antibiotics

Nystatin belongs to the group of the polyene
macrolide compounds, which are characterized by having 3
to 8 conjugated double bonds in their macrolactone ring.
Nystatin itself is a tetraene, having 4 conjugated
double bonds between C20 and C27. There is also a set
of 2 conjugated double bonds on the nystatin molecule,
between C30 and C33, which is separated from the set of
4 conjugated double bonds by C28-C29 (see Fig. 1). From
the computer-assisted analysis of the NysC PKS it became
apparent that the ER domain in module 5 in this protein
is responsible for reduction of the double bond between
C28 and C29. Thus, by inactivating this particular
domain, it is theoretically possible to obtain a
compound with a double bond between C28 and C29, thus
joining two sets of conjugated double bonds in the
nystatin molecule, and creating a heptaene macrolide
compound.

To inactivate the ER domain in module 5 of NysC,
the method of "in-frame" deletion within the *nysC* gene
was chosen. The construction of the vector pERD4.2 for
gene replacement in the NysC-encoding genomic region of

- 69 -

S. noursei was as follows:

Inactivation of ER domain in module 5 of NysC

PCR-assisted amplification of the 394 bp DNA
fragment representing the coding sequence for the C-
terminal part of the ER domain in module 5 of NysC, and
the coding sequence for the N-terminal part of the KR
domain in module 5 (nt 32174 - 32559), was performed.
The oligonucleotide primer ERD1 (5'-
GTTGGTACCCCACTCCCGGTCCGCAC-3', sense) (SEQ ID NO. 25) was
selected from the nucleotide sequence of *nysC* gene
comprising nt 32174-32190 (SEQ ID NO. 1) with additional
nucleotides on the 5' end in order to create a *KpnI*
restriction enzyme cleavage site. The oligonucleotide
primer ERD2 (5'-CCAGCCGCATGCACCACC-3', antisense (SEQ ID
NO. 26)) was selected from the *nysC* coding DNA sequence,
and comprised the DNA segment between nt 32559-32542
(SEQ ID NO. 1) containing a *SphI* restriction enzyme
cleavage site. The resulting PCR fragment was digested
with *KpnI* and *SphI*, and ligated together with the 1828
bp *BamHI/KpnI* DNA fragment (nt 29224-31052 of SEQ ID NO.
1), and the 1273 bp *SphI/EcoRI* DNA fragment (nt 32548-
33821 of SEQ ID NO. 1) into the *EcoRI/BamHI* - digested
pGEM3Zf(-) vector. The ligation mixture was transformed
into the *E.coli* DH5α, and recombinant plasmid of 6.7 kb
designated pERD4.1 was recovered from one of the
transformants. The latter contained a hybrid DNA
fragment representing nt 29224-33821 of the SEQ ID NO. 1
DNA sequence with internal deletion between nt 31052 and
nt 32174 of the *nysC* coding region. This deletion
eliminated the coding regions (aa 4837 to 5208 of *nysC*)
for the part of the DH-ER interdomain linker and C-
terminal part of the ER domain containing a putative
NADP(H) binding site.

To construct the vector for inactivation of NysC
ER4 domain in *S.noursei*, the recombinant DNA fragment
was excised with *EcoRI* and *HindIII* restriction enzymes

- 70 -

from pERD4.1 and ligated with the 3.0 kb EcoRI/HindIII
DNA fragment from pSOK201, and the ligation mixture was
transformed into the E.coli DH5α. Plasmid pERD4.2 of
6.5 kb was isolated from one of the transformants and
5   used to perform the gene replacement procedure in
S.noursei ATCC11455 (see below). The recombinant
S.noursei strains selected after this procedure were
designated ERD44 and ERD48.

The latter plasmid was introduced into S. noursei
10   ATCC11455 by intergeneric conjugation, and one
transconjugant, S. noursei (pERD4.2), was chosen for
further manipulations. After the correct mode of
integration of pERD4.2 into the genome of the S. noursei
(pERD4.2) was confirmed by Southern blot analysis,
15   selection for the second crossover event was carried out
as described in Sekurova et al., FEMS Microbiol Lett,
v.177: 297-304, 1999 (and see below).

Gene replacement procedure
20   This method is carried out as described by Sekurova
et al., 1999, supra.

The plasmid constructed for gene replacement as
described above was introduced into S.noursei by
conjugation from the E.coli ET 12567 (pUZ8002). One of
25   the clones carrying the plasmid integrated into the
chromosome via homologous recombination was subjected to
three rounds of sporulation on antibiotic-free ISP2 agar
medium, and the progeny after the third round was tested
for the loss of antibiotic resistance marker. Southern
30   blot analysis of the total DNA isolated from several
antibiotic-sensitive strains with an appropriate probe
was used to confirm the desired mutation.

Of 8 colonies, which had lost the selection marker,
and thus undergone a second crossover event, 4 were
35   shown by Southern blot analysis to have reverted to the
wild-type genotype. Two strains were shown to contain a
large deletion apparently eliminating a substantial

portion of the nystatin gene cluster, while two other
mutants contained either the desired 1116 bp deletion
(ERD44), or a deletion which was somewhat larger than
expected (ERD48). Analysis of the polyene antibiotic
5   production by the strains ERD44 and ERD48 revealed that
they do not produce nystatin. Instead, ERD44 was shown
to produce a polyene compound having UV spectrum peaks
characteristic for heptaens (Figure 6), as expected.
Surprisingly, the ERD48 mutant was shown to produce
10   hexaenic polyenes (according to spectroscopic analysis),
which would be consistent with "in-frame" deletion of
the complete module 5 from the NysC protein. In order to
investigate the event which occurred in the ERD48 mutant
in more detail, the DNA fragment was PCR-amplified from
15   the genome of *S. noursei* ERD48 mutant, which would
encompass the putative product of such deletion (see
below).

PCR-assisted amplification and cloning of the PKS-
20   ***encoding DNA fragment from the S.noursei ERD48 genome***
The PCR reaction aimed at amplification of part of
the mutant *nysC* gene in *S.noursei* ERD48 was carried out
with oligonucleotide primers KR48.1 (sense, 5'-CCG CGT
CGG ATC CGC CGA C-3') (SEQ ID NO: 27) and KR48.2
25   (antisense, 5'-AGC CTT CGA ATT CGG CGC C-3') (SEQ ID NO:
28) which corresponded to the nt 24744-24760 and
nt33818-33833, respectively, of the DNA sequence in SEQ
ID NO. 1. The 50 μl PCR mixture contained 0.1 μg of
S.noursei ERD48 total DNA, 25 pm of each KR48.1 and
30   KR48.2 oligonucleotide primers, dNTPs (final
concentration 350μm), 1xPCR buffer from Expand High
Fidelity PCR System (Boehringer Mannheim) and 1.5 U of
the DNA polymerase mixture from the same system. The
PCR was performed on the Perkin Elmer GeneAmp PCR System
35   2400 with the following program: 1 cycle of denaturation
at 96°C (4 min), 35 cycles of denaturation/annealing/
synthesis at 94°C (45 sec) and 70°C (10 min) and 1 full

- 72 -

cycle of final annealing/extension at 72°C (10 min).
The 2.7 kb DNA fragment obtained with this procedure was
digested with *EcoRI* and *BamHI* restriction enzymes, and
ligated with pGEM3Zf(-) vector DNA digested with the
5   same enzymes. The ligation mixture was introduced into
E.coli DH5α by transformation, the plasmid pKR48 of
5.9kb was isolated from one of the transformants and
subjected to DNA sequence analysis.

The DNA sequence of the insert in pKR48 is present
10   in SEQ ID NO. 29, (identified herein as ERD48 seq). The
translation product is shown in SEQ ID NO. 30 and is a
899 aa protein - the molecule features of SEQ ID NOs 29
and 30 respectively are shown below:

15   <u>SEQ ID NO: 29 (DNA; ERD48.seq)</u>

Start	End	Name	Description
1	254	DH4	DH4 domain coding region, C-terminal
1170	1913	KR4/5	hybrid ketoreductase domain, module 4/5
2010	2231	ACP5	ACP5 domain coding region
2295	2700	KS5	KS5 domain coding region

<u>SEQ ID NO: 30 (AA; translation product)</u>

Start	End	Name	Description
1	84	DH4	DH4 domain module 4, C-terminus
390	637	KR4/5	hybrid KR domain
670	743	ACP5	ACP domain, module 5
764	899	KS5	KS domain, module 5, N-terminus

30   The cloning and DNA sequencing of this 2700 bp
fragment confirmed that it encodes a part of a hybrid
PKS module, which would be consistent with recombination

- 73 -

between DNA sequences encoding highly homologous KR
domains in modules 4 and 5 of NysC. This recombination
apparently has led to deletion of DNA sequences encoding
C-terminal parts of KR and ACP domains of module 4, and
KS, AT, DH, ER domains, and N-terminal part of the KR
domain of module 5, thus resulting in the loss of one
complete module from NysC PKS.

Preliminary analysis of the compounds produced by
the *S. noursei* ERD44 and ERD48 was carried out. It was
shown that a heptaenic compound produced by the ERD44
mutant has high antifungal activity against *Candida
albicans*, an organism used in tests for antifungal
activity. At this point it is not possible to
accurately assay the activity of this compound
(tentatively named S44), because it is not yet properly
purified, and its exact concentration is difficult to
estimate. However, some rough estimates based on the UV
absorbance at the wave lengths characteristic for
nystatin, S44, and amphotericin (a heptaenic macrolide),
suggest that S44 might have 4-5 times higher antifungal
activity compared to nystatin. HPLC analysis of the
compounds produced by the ERD48 mutant suggests that at
least 5 hexaenic macrolides with different retention
times are produced by this strain (mixture called S48).
This probably reflects the different states of
modifications of the macrolactone ring by i.e.
glycosylation at C19, hydroxylation at C10, or oxidation
of the methyl group at C16. This could have been
expected, since reduction of the macrolactone ring size
most probably leads to the lower affinity of the
modifying enzymes towards the new substrate.

Antifungal activity of the S48 mixture was tested,
and found to constitute approx. 10% of nystatin
activity. It seems probable that only one of the
compounds in the S48 mixture produced by ERD48, which is
fully decorated by the ring-modifying enzymes, is
responsible for the antifungal activity detected. Thus,

- 74 -

relative antifungal activity of this compound is
impossible to assess prior to its purification. Some of
the hexane antibiotics are known to have antibacterial,
as well as antifungal activity (Ciftci, et al., J.
5   Antibiot, v. 37: 876-884, 1984). It is thus possible
that hexaenic compounds produced by the ERD48 mutant can
be used for production of antibacterial agents. The
changes in the NysC protein leading to production of new
polyene compounds in ERD44 and ERD48 mutants, along with
10   the predicted structures of their macrolactone rings are
presented in Fig. 5.

Inactivation of DH8 domain in module 8 of NysC
The possibility of genetically manipulating the
15   nystatin PKS was further exemplified by inactivation of
the DH domain in module 8 of NysC. The plasmid pNPR1.1
for gene replacement within *nysC* gene, which would
result in in-frame deletion of the DNA region encoding
DH8 domain was constructed as below:
20   The 3989 bp *KpnI/BclI* DNA fragment (nt 43004-46993
of the region 1 DNA sequence (SEQ ID NO. 1) and the 2409
bp *BamHI/EcoRI* DNA fragment (nt 47680-50089 of the same)
were excised from the DNA of recombinant phage N1 and
ligated with vector pGEM3Zf(-) DNA digested with *EcoRI*
25   and *KpnI*. The ligation mixture was transformed into
E.coli DH5a, and recombinant plasmid pGEM-NPR1 was
isolated from one of the transformants. The latter
contained the hybrid DNA fragment representing the nt
43004-50089 of the region 1 DNA sequence (SEQ ID NO:1)
30   with the internal deletion between nt 46993 and nt
47680. This deletion eliminated the DNA region encoding
the aa 10150 to 10378 of the NysC polypeptide, thus
affecting the DH8 domain in module 8, and DH8-KR
interdomain linker in module 8. To construct the vector
35   for inactivation of the NysC DH8 domain in *S.noursei*,
the recombinant DNA fragment was excised with *EcoRI* and
*Hind*III restriction enzymes from pGEM-NPR1 ligated with

- 75 -

the 3.0 kb EcoRI/HindIII DNA fragment from pSOK201 (see
Fig. 3 and Table 3) and the ligation mixture was
transformed into the E.coli DH5α. Plasmid pNPR1.1 of
9.4 kb was isolated from one of the transformants and

5   used to perform the gene replacement procedure in
S.noursei ATCC11455 according to Sekurova et al., 1999,
supra.

         The recombinant S.noursei strain selected after
this procedure was designated NPR1.1 The S. noursei

10   NPR1.1 recombinant strain was shown by Southern blot
analysis to contain the desired deletion in the
DH8-coding sequence of nysC. Analysis of the secondary
metabolites in the culture extracts of the S. noursei
NPR1 recombinant strain by thin layer chromatography

15   (TLC) revealed the presence of presumed macrolide
compounds. The relative mobility of these compounds
differed from nystatin, and no UV spectra characteristic
for nystatin could be detected in the extracts. It was
suggested, that in the new molecule(s) produced by the

20   NPR1 recombinant a set of 4 double bonds on the nystatin
aglycone has been disturbed, and that the macrolactone
ring now contains a hydroxy group attached at C23 (Table
5). No attempts to purify the compound(s) from NPR1
were made, as the bioassay against Candida albicans made

25   with the NPR1 culture extracts showed very low
antifungal activity. However, the NPR1 mutant can be
potentially useful for further manipulations with the
nystatin PKS.

30   *Inactivation of KR domain in module 7 (NysC)*
         The 4404 bp DNA fragment was excised with EcoRI and
SmaI restriction enzymes from the DNA of recombinant
phage N1. The EcoRI site is situated in the polylinker
of phage N1 to the left of the S.noursei DNA insert

35   starting at nt 38398 of the region 1 DNA sequence (SEQ
ID NO. 1), while the SmaI site corresponds to nt 42802
of the region 1 DNA sequence. The 3303 bp DNA fragment

- 76 -

was excised with *Sma*I and *Bam*HI restriction enzymes from
the DNA of the same recombinant phage. The *Sma*I and
*Bam*HI restriction sites are situated at nt 43099 and nt
46402 respectively, of the region 1 DNA sequence (SEQ ID
5   NO. 1). Both DNA fragments were ligated with the
pGEM3Zf(-) DNA digested with restriction enzymes *Eco*RI
and *Bam*HI, and the ligation mixture was transformed in
Escherichia coli DH5α. The plasmid pGEM-NPR2 of 10.7 kb
was recovered from one of the transformants, which
10   contained a recombinant DNA fragment which represented
nt 38398-46402 of the region 1 DNA sequence (SEQ ID NO.
1) with the internal deletion between nt 42802 and nt
43099. This deletion results in elimination of the DNA
region encoding aa 8753 to 8851 of the NysC protein
15   encompassing a putative NADP(H) binding site in the KR
domain of module 7. To construct the vector for
inactivation of NysC KR7 domain in *S.noursei*, the
recombinant DNA fragment was excised from pGEM-NPR2 with
*Eco*RI and *Hind*III restriction enzymes, and ligated with
20   the 3.0 kb *Eco*RI/*Hind*III DNA fragment from pSOK201, and
the ligation mixture was transformed into the *E.coli*
DH5α. Plasmid pNPR2 of 10.7 kb was isolated from one of
the transformants and used to perform the gene
replacement procedure in *S.noursei* ATCC11455 (according
25   to Sekurova et al., 1999, supra). The recombinant
S.noursei strain selected after this procedure was
designated NPR2.1.
         Analysis of the secondary metabolites in the
culture extracts of the *S. noursei* NPR2.1 recombinant
30   strain by TLC revealed the presence of presumed
macrolide compounds. The relative mobility of these
compounds differed from nystatin, and from the
metabolites produced by NPR1 mutant. No UV spectra
characteristic for nystatin could be detected in the
35   extracts. It is suggested, that in the new molecule(s)
produced by the NPR2.1 recombinant a set of 4 double
bonds on the nystatin aglycone has been disturbed, and

- 77 -

the macrolactone ring now contains a keto group at C25
(Table 5). No attempts at purification of the
compound(s) from NPR2.1 were made, and no bioassays for
antifungal activity with the NPR2.1 culture extracts

5   were performed. This mutant has utility not only by
virtue of the metabolites it produces, but also for
further manipulation with nystatin PKS.
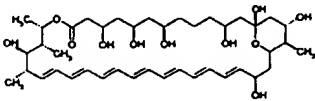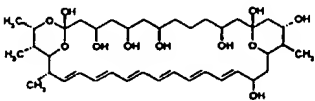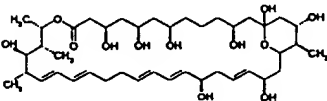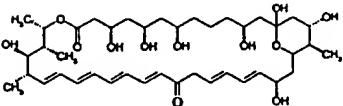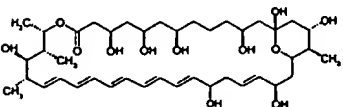
Inactivation of the ER domain in module 5 of the mutated
10   *NysC protein NysC_DH8)*

       To introduce the second mutation into the NysC
protein with inactivated DH domain in module 8
(NysC_DH7), the plasmid pERD4.2 was introduced into the
S.noursei mutant NPR1.1 and the gene replacement

15   procedure was carried out as described in Sekurova et
al., 1999, supra. This yielded recombinant S.noursei
strain ERDH9 with mutations in both ER5 and DH8 coding
sequences of nysC. The combination of these two
mutations presumably leads to biosynthesis of the

20   pentaenic nystatin derivative with a hydroxy group at
C23 (Table 5). Preliminary analysis of the ERDH9
culture extracts confirmed that a polyene compound(s) is
being produced by this strain although in quantities
making identification of its true UV spectrum difficult.

25   The preliminary data also show that this compound is
preferentially accumulated in the culture supernatant,
while nystatin produced by the wild-type S. noursei
remains mostly associated with mycelium. This was
consistent with the hypothesis that an additional

30   hydroxy group on the nystatin molecule is responsible
for increased water solubility of the compound(s). No
attempts at purification of this new compound(s) were
made, and no bioassays were performed.

       To date, the Inventors have been unable to confirm
35   the structure of the products of mutants NPR1.1, NPR2.1
and ERDH9.

Table 5. Predicted structures of nystatin derivatives produced via genetic engineering of *nysC* gene in *S. noursei* (see Example 2 for details)

Mutant	Expected structure (polyketide moiety only)	UV spectrum, nm (DMSO)	Activity
ATCC11455	nystatin (Fig. 1)	299, 312, 327	normal
ERD44		375, 395, 419, 444	high
ERD48		336, 352, 370, 391	low?
NPR1		-	low
NPR2.1		-	?
ERDH9		?	?

- 79 -

Example 3 - Manipulation of the regulatory genes leading
to increased production of nystatin

Expression of nysR1 under the control of PermE* promoter
in S. noursei ATCC11455

To further confirm that *nysR1* gene encodes a
transcriptional activator of the nystatin biosynthesis
genes, the latter was expressed in *S. noursei* ATCC11455
under the control of the PermE* promoter (Bibb et al.,
Mol. Microbiol., v. 14: 433-545, 1994). First, the
plasmid pSOK804 for stable and efficient integration
into the *S. noursei* ATCC11455 genome was constructed
(Fig. 2). This was made by ligating together the 3.0 kb
*Sph*I/*Hind*III DNA fragment from pSET152 (Bierman et al.,
1992, supra) and 2.3 kb *Sph*I/*Hind*III fragment from
bacteriophage VWB carrying functions necessary for site-
specific integration (Van Mellaert et al., 1998, supra).
Conjugation of pSOK804 from E. coli ET12567 (pUZ8002)
into *S. noursei* ATCC1455 demonstrated that this plasmid
integrates specifically into one site of the *S. noursei*
genome at a frequency of $3-10^{-6}$.

To clone the *nysR1* gene under PermE* promoter in
pSOK804, the following procedure was employed. The N-
terminal part of *nysR1* was PCR-amplified from the
recombinant phage N1 DNA template with the
oligonucleotide primers NR1.1 (sense) 5'-
CGCCGCATGCTGTTCTCACCCCACGT-3' (SEQ ID NO: 31), and NR1.2
(antisense): 5'-GGCGCGACCCGGTTCGGCCT-3' (SEQ ID NO: 32).
The oligonucleotide primer NR1.1 sequence corresponded
to the nt 51376-51391 of SEQ ID NO: 1 with addition of
nucleotides CGCC<u>GCATGC</u> at the 5' end to create a site
for *Sph*I endonuclease. The oligonucleotide primer NR1.2
corresponded to the sequence complementary to nt 51964-
51982 of SEQ ID NO: 1, and encompassed a restriction
site for *Age*I endonuclease. A 0.6 kb fragment was PCR-
amplified using the phage N1 DNA as a template under
conditions decribed in Example 1, and digested with *Sph*I

- 80 -

and *AgeI*. The digested 0.6 kb fragment was ligated
together with a 2.8 kb *AgeI/EcoRI* DNA fragment from the
phage N1 insert (*EcoRI* site originating from the phage's
polylinker) into pGEM7Zf(-) vector digested with *SphI*
and *EcoRI*. The 2.8 kb fragment used in this ligation
corresponded to nt 51971-54814 of SEQ ID NO: 1, and
encompassed the C-terminal part of *nysR1* and N-terminal
part of *nysR2* (encoding the first 162 aa of NysR2). The
recombinant plasmid obtained as a result of this
ligation was designated pNRE1. The 3.5 kb *SphI/HindIII*
DNA fragment from pNRE1 was ligated together with the
0.3 kb *EcoRI/SphI* fragment from pGEM7ermELi (see Table
3) containing PermE* promoter into pSOK804 vector
digested with *EcoRI/HindIII*. The resulting plasmid,
pNRE2, was introduced into the *S. noursei* ATCC11455 by
conjugation (see Example 1) yielding recombinant strain
S. noursei (pNRE2). Analysis of the nystatin production
by the latter strain in shake-flasks with reduced
glucose medium revealed that it produces 50% more
nystatin compared to the wild-type strain, most likely
due to overexpression of the *nysR1* gene from the PermE*
promoter. It therefore appears that *nysR1* gene encodes
a positive activator that may be used for enhancing the
production of nystatin and its derivatives in *S. noursei*
strains.

Partial deletion of nysR5 gene in the S. noursei
ATCC11455

       To confirm the function of the *nysR5* gene predicted
through analysis of its coding DNA sequence (see Example
1) in the regulation of nystatin biosynthesis, a
specific mutation in *S. noursei* ATCC11455 genome was
introduced. A DNA fragment from the *S. noursei*
ACTT11455 genome encompassing nt 62037-63360 of the
nucleotide sequence reported here was amplified by PCR
with the primers NR5D1 (5'-GCGAGCGGCCGCTTCACCCCGCAACTCA-
3') (SEQ ID NO: 33) and NR5D2 (5'-

- 81 -

CGCGAAGCTTGGCCGACTGCTCGACGTC-3') (SEQ ID NO: 34). The
conditions for PCR were as described above. The 1341 bp
PCR product was digested with *NotI* and *HindIII*, and
ligated with the 1688 bp *EcoRI/NotI* DNA fragment (nt
5   60301-61989) and 3.0 kb *EcoRI/HindIII* fragment from
pSOK201. The resulting plasmid, pNR5D, contained the *S.
noursei* DNA fragment with a 43 bp deletion in the coding
region of *nysR5* gene. This deletion creates a frame-
shift mutation within the *nysR5* coding region,
10   subsequently leading to truncation of its product. As a
result of such truncation, 165 C-terminal amino acids of
NysR5 are eliminated, and replaced with 14 amino acids
encoded by another reading frame (and thus unrelated to
NysR5). The pNR5D plasmid was used to perform a gene
15   replacement procedure in *S. noursei* ATCC11455 as
described in Sekurova et al., 1999, supra.

   The mutation introduced through gene replacement
led to a 5-15% increase in nystatin production by the
resulting recombinant strain NR5D, compared to the wild-
20   type *S. noursei*. Subtle but reproducible positive
effect of NysR5 C-terminal deletion on nystatin
biosynthesis correlates well with the putative repressor
function assigned to this protein based on computer
analysis (see Example 1). Since the deletion introduced
25   in the *nysR5* gene does not eliminate the N-terminally
located putative helix-turn-helix motif identified in
this protein, the residual repressor activity of the
truncated NysR5 polypeptide could account for the
relatively small effect of this mutation on nystatin
30   production. Nevertheless, this result confirms the
usefulness of introducing mutations in the repressor-
encoding gene as further means for enhancing the
production of nystatin and its derivatives.

35

- 82 -

Example 4 - Completion of the sequencing of the nystatin
biosynthesis gene cluster

    The DNA sequence spanning the gap between SEQ ID
No. 1 and SEQ ID No. 2 was determined for both DNA
strands on the overlapping inserts in recombinant phages
N20, N32, N95, N98, and N99 covering this region.
    Procedures used for sequencing, probe generation
and screening were as described in Example 1. The
probes used for library screening, and the relevant
recombinant phages discovered were as follows:

        PKS 72   N20, N32
        L44ES3.5   N90
        L76SN0.5   N95
        L20S0.64   N98, N99

Description of the probes:
        6.   L44ES3.5 probe. The 3.5 kb DNA fragment
isolated from the phage N44 with restriction enzymes
EcoRI and SalI, was used as a L44ES3.5 probe.
        7.   L76SN0.5 probe. The 0.5 kb DNA fragment
isolated from the recombinant phage N76 with restriction
enzymes SacI and NotI, was used as a L76SN0.5 probe.
        8.   L20S0.64 probe. The 0.64 kb DNA fragment
isolated from the plasmid pL20EB3.7 with the restriction
enzyme SalI, was used as a L20S0.64 probe.
    New sequence information resulted in identification
of complete nysI and nysDII genes, as well as the new
genes nysJ, nysK, nysL, nysM, and nysN (see Figure 7).
According to the new information, the NysI protein of
9477 aa (SEQ ID No. 37) represents a PKS composed of six
modules, responsible for the elongation steps 9 to 14 of
the nystatin polyketide backbone biosynthesis. All
these modules contain KS, AT, DH, KR and ACP domains.
The presence of a mAT domain in module 11 is consistent
with incorporation of methylmalonyl-CoA extender at this

- 83 -

elongation step. The DH domains in modules 10, 11, 12, and 14 seem to be inactive due the absence of the active site motif H(X$_3$)G(X$_4$)P. The KR domain in module 13 of NysI lacks the conserved motif aSRrG, and thus appears

5   to be inactive. The latter feature, together with an inactive DH domain in module 11, most probably account for the presence of a six-membered ketalic ring (between C13 and C17) on the nystatin molecule.

The *nysJ* gene encoding a PKS, is located downstream

10   of *nysI*, and is transcribed in the same direction. As judged from the organisation of modules in NysJ (SEQ ID No. 38), the latter is required for the elongation steps 15 to 17 in the nystatin macrolactone ring assembly. The DH domains in modules 16 and 17 within NysJ seem to·

15   be inactive, and the ER domain localized in module 15 is most probably responsible for the reduction of a double bond between C8 and C9.

The last, 18th module in the nystatin PKS system is represented by the NysK protein (SEQ ID No. 39) encoded

20   by the *nysK* found downstream of *nysJ*. The NysK protein of 2066 aa is composed of KS, AT, inactive DM, ACP and TE domains. The NysK protein lacks a KR domain, and contains apparently inactive DH domain. A TE domain was identified at the C-terminus of NysK, suggesting that in

25   addition to the condensation of the last extender unit, this protein also participates in the release of the mature nystatin polyketide chain from the PKS complex.

The gene organisation of the cluster is shown in Figure 9 and Figure 8 sets out the proposed involvement

30   of the various proteins encoded in the nystatin biosynthetic pathway.

To confirm the involvement of *nysI*, and *nysJ* in nystatin biosynthesis, these genes were disrupted in *S. noursei* via homologous recombination using the

35   conjugative suicide vectors pKNI1, and pKNJ1. The construction of the above vectors for disruption experiments was as follows. A 3.8 kb *EcoRI/BamHI*-

- 84 -

fragment internal for the *nysI*, and corresponding to
nucleotides 23711-27541 of SEQ ID No. 35, was excised
from N64 DNA and ligated into the corresponding sites of
plasmid pGEM3Zf(-), resulting in plasmid pL64EB3.8. The

5   *S. noursei* fragment cloned into pL64EB3.8 was excised
from this plasmid with restriction enzymes *Eco*RI and
*Hind*III, and ligated with the 3.0 kb *Eco*RI/*Hind*III-
fragment from the vector pSOK201, resulting in plasmid
pKNI1.

10   A 3.7 kb *Eco*RI/*Bam*HI-fragment internal for *nysJ*,
and corresponding to nucleotides 43287-46992 of SEQ ID
No. 35 was excised from phage N20 DNA, and ligated into
the corresponding sites of plasmid pGEM3Zf(-), resulting
in plasmid pL20EB3.7. The *S. noursei* fragment cloned

15   into pL20EB3.7 was excised from this plasmid with
restriction enzymes *Eco*RI and *Hind*III, and ligated with
the 3.0 kb *Eco*RI/*Hind*III-fragment from the vector
pSOK201, resulting in plasmid pKNJ1. Both pKNI1 and
pKNJ1 constructs were transformed into *E. coli* ET 12567

20   (pUZ8002), and further transferred into *S. noursei* ATCC
11455 by conjugation, as described in Zotchev *et al.*,
(2000) Microbiology 146: 611-619. No nystatin
production was detected in either of the pKNI1 and pKNJ1
disruption mutants, thus confirming the requirement of

25   the identified PKS's for nystatin biosynthesis.

Three genes encoding proteins presumably involved
in modification of the nystatin molecule were identified
between *nysK* and *nysDII*. Both the *nysL* and *nysN* genes
encode P450 monooxygenases of 394 aa and 398 aa,

30   respectively (SEQ ID Nos. 40 and 42 respectively), that
are probably responsible for hydroxylation of the
nystatin polyketide moiety at C10, and oxidation of the
methyl group at C16. Which protein is responsible for
which reaction is not yet clear, and additional

35   experiments are required for exact placement of NysL and
NysN in the nystatin biosynthetic pathway. The *nysM*
gene apparently encodes a ferredoxin of 64 aa (SEQ ID

No. 41), which presumably constitutes a part of one or both P450 monooxygenase systems, and serves as an electron donor [O'Keefe, D.P. & Harder, P.A. (1991). Occurrence and biological function of cytochrome P450 monooxygenases in the actinomycetes. *Mol. Microbiol.* 5, 2099-2105].

The DNA sequences extending the region depicted on Figure 7 (SEQ ID No. 35) approximately 10 kb to the left (recombinant phage N90), and approximately 5 kb to the right (recombinant phage N69) were determined. No genes with plausible functions in the nystatin biosynthesis were found, suggesting that the entire nystatin gene cluster had been identified.

Thus, based on the complete sequence information for the nystatin biosynthetic gene cluster, the following genes are identified and their roles described as follows (see also Figures 8 and 9):

Table 6. Genes identified within the nystatin biosynthetic gene cluster of *S. noursei* ATCC 11455

Designation	Product	Putative function
nysF putative	4'-phosphopantheteine transferase	post-translational PKS modification
nysG	ABC transporter	efflux of nystatin
nysH	ABC transporter	efflux of nystatin
nysD3	GDP-mannose-4,6-dehydratase	mycosamine biosynthesis
nysI	PKS type I	nystatin PKS (modules 9-14)
nysJ	PKS type I	nystatin PKS (modules 15-17)
nysK	PKS type I	nystatin PKS (module 18 + TE)
nysL	P450 monooxygenase	hydroxylation at

- 86 -

			C-10
	nysM	ferredoxin	electron transfer in P450 system
5	*nysN*	P450 monooxygenase	oxidation of methyl group at C-16
	nysD2	aminortansferase	mycosamine biosynthesis
	nysD1	glycosyltransferase	attachment of mycosamine
10	*nysA*	PKS type I	nystatin PKS (loading module)
	nysB	PKS type I	nystatin PKS (modules 1 and 2)
15	*nysC*	PKS type I	nystatin PKS (modules 3-8)
	nysE	thioesterase	release of polyketide chain from PKS
20	*nysR1*	transcriptional activator	regulation of nystatin production
	nysR2	transcriptional activator	regulation of nystatin production
	nysR3	transcriptional activator	regulation of nystatin production
25	*nysR4*	transcriptional activator	regulation
	nysR5	transcriptional repressor	regulation
30	ORF2	transcriptional activator	regulation
	ORF1	peptidase	peptide metabolism

- 87 -

Example 5 - Manipulation of nystatin biosynthetic genes:

Predictive examples presented below provide
guidelines for the rational genetic manipulations of the
nystatin biosynthetic genes aimed at specific chemical
changes in the nystatin molecule. These manipulations
are based on the current understanding of structure-
function relashionship of the polyene antibiotics the
number of conjugated double bonds and the presence of two
ionisable groups (exocyclic carboxyl and aminogroup
belonging to the deoxysugar moiety mycosamine).

*Changing the number and positions of conjugated double
bonds.*

The conjugated double bonds within the nystatin
macrolactone ring are formed as a result of two reductive
steps performed by a PKS modules with ketoreductase (KR),
and dehydratase (DH) activities. Further reduction of
the double bond can be brought about by introducing a
enoylreductase (ER) activity in such PKS modules. This
shall result in the formation of a completely saturated
bond instead of a double bond at a specific step of
nystatin biosynthesis. The following manipulations can
be proposed (compounds that theoretically can be produced
as a result of these manipulations are presented on
Figure 10):
-     insertion of ER domain into module 3 (1)
-     insertion of ER domain into module 4 (2);
-     simultaneous inactivation of the ER domain in
      module 5 and insertion of the ER domain into module
      3 (3);
-     simultaneous inactivation of the ER domain in
      module 5 and insertion of the ER domain into module
      4 (4);
-     simultaneous inactivation of the ER domain in
      module 5 and insertion of the ER domain into module

- 88 -

7 (5);

- simultaneous inactivation of the ER domain in
  module 5 and insertion of the ER domain into module
  8 (6);

5   - simultaneous inactivation of the ER domain in
         module 5 and insertion of the ER domain into module
         9 (7);

       - simultaneous inactivation of the ER domain in
         module 5 and insertion of the ER domains into

10   modules 8 and 9 (8);

       - simultaneous inactivation of the ER domain in
         module 5 and insertion of the ER domains into
         modules 7 and 8 (9);

15   These manipulations can be performed using the
       techniques for gene replacement described for *S. noursei*
       (Sekurova et al., FEMS Microbiol Lett, **177**: 297-304,
       1999). The materials for manipulations can be provided
       by the nystatin gene cluster itself, or other PKS gene

20   clusters. The latter could be preferential from the
       point of view of genetic stability of the recombinant
       strains.

       Removal/repositioning of the COOH group

25

       Exocyclic carboxyl function of the polyene
       antibiotics is believed to play a crucial role in
       selective toxicity of these compounds. More
       specifically, the inter- and intramolecular interactions

30   between the ionizable exocyclic carboxyl and amino group
       of micosamine moiety seem to be of particular importance
       (Resat, H., Sungur, F.A., Baginski, M., Borowski, E.,
       Aviyente, V. 2000. Conformational properties of
       amphotericin B amide derivatives - impact on selective

35   toxicity. Journal of Computer-Aided Molecular Design, v.
       14, p 689-703). It is theoretically possible to either
       remove the exocyclic carboxyl from, or reposition it on

- 89 -

the nystatin moleculevia manipulation of the nystatin
biosynthetic genes. Nystatin derivatives produced via
such manipulations could be useful on their own, or serve
as substrates for further chemical modifications.

5   The following manipulations are proposed (resulting
compounds are represented on Figure 11):

- replacement of methylmalonyl-specific
  acetyltransferase (AT) domain in module 11 of the
  nystatin PKS with malonyl-specific AT domain (10);

10 - replacement of malonyl-specific AT domain in module
  12 with methylmalonyl-specific AT domain with
  simultaneous replacement of methylmalonyl-specific
  AT domain in module 11 with malonyl-specific AT
  domain (11);

15 - replacement of malonyl-specific AT domain in module
  10 with methylmalonyl-specific AT domain with
  simultaneous replacement of methylmalonyl-specific
  AT domain in module 11 with malonyl-specific AT
  domain (12);

20 - inactivation of P450 monooxygenase-encoding genes
  nysL or *nysN* (whichever is found to be responsible
  for oxygenation of the methyl group at C-16 on the
  nystatin molecule) (13).

25   It shall be noted that specificity of the P450
monooxygenase responsible for the appearance of the
exocyclic carboxyl function can be engineered so that it
fulfills its function on the new substrates. Such
methods as site-specific or random mutagenesis along with
30 error-prone PCR and DNA shuffling might prove useful for
this purpose.

*1.3. Introduction of additional hydroxyl functions
(increasing water solubility).*

35

      Polyene antibiotics are very poorly soluble in
water mostly due to a highly hydrophobic set(s) of

- 90 -

conjugated double bonds. Increasing water solubility can
be an advantage in certain cases, as it expands
pharmacologicxal properties of the drug (Golenser, J.,
Frankenburg, S., Ehrenfreund, T. & Domb, A.J. 1999.
5   Efficacious treatment of experimantal leishmaniasis with
amphotericin B-arabinogalactan water-soluble derivatives.
Antimicrob Agents Chemother., v. 43: 2209-2214). To
increase the water solubility of nystatin, we suggest to
introduce additional hydroxyl functions (hydrophilic) to
10   the nystatin molecule. The following modifications of
the nystatin biosynthetic genes can lead to the desired
effect (resulting compounds depicted on Figure 12):
-    inactivation of dehydratase (DH) domain in module 3
of the nystatin PKS (**14**);
15   -   inactivation of DH domain in module 4 (**15**);
-    inactivation of DH domain in module 3 with
simultaneous inactivation of ER domain in module 5
(**16**);
-    inactivation of DH domain in module 4 with
20   simultaneous inactivation of ER domain in module 5
(**17**) ;
-    inactivation of DH domain in module 7 with
simultaneous inactivation of ER domain in module 5
(**18**);
25   -   inactivation of DH domain in module 8 with
simultaneous inactivation of ER domain in module 5
(**19**);
-    inactivation of DH domain in module 9 with
simultaneous inactivation of ER domain in module 5
30   (**20**);

**Extension, truncation or rebuilding of the macrolactone
ring.**

35   Novel derivatives of a polyene antibiotic nystatin
can be obtained also through truncation of the nystatin
PKS, leading to derivatives with a smaller macrolactone

ring (as exemplified in example 2 for the ERD48 mutant).
This can be achieved through deletion of one or more
modules from the nystatin PKS. Such truncations can lead
to production of polyketides with 36- to 6-membered
5   lactone rings that potentially can be useful for further
modifications and synthesis of novel pharmaceuticals.
Extension of the nystatin macrolactone ring can be
achieved through insertion of additional modules into the
nystatin PKS. Such manipulations can also lead to
10   production of the lead compounds for pharmaceutical
applications.

    The nystatin molecule can be completely rebuilt by
the way of shuffling the PKS modules between the
nystatin, or other PKSs so that completely new
15   derivatives are produced. In this respect, the method
disclosed in the patent WO 00/77181 can prove useful for
making the recombinant DNA constructs serving this
purpose.

    Finally, the nystatin biosynthetic genes can prove
20   useful for manipulation of other macrolide antibiotic
biosynthetic pathways. Both PKS and modification enzymes
can prove useful for such purposes. It seems likely that
nystatin biosynthetic genes will be most useful for
manipulation of other polyene antibiotic biosynthetic
25   clusters, such as the one for pimaricin (Aparicio, J.F.,
Fouces, R., Mendes, M.V., Olivera, N., Martin, J.F. 2000.
A complex multienzyme system encoded by five polyketide
synthase genes is involved in the biosynthesis of the 26-
membered polyene macrolide pimaricin in *Streptomyces*
30   *natalensis*. Chem Biol, v. 7: 895-905). High degree of
similarity on the protein level between the nystatin and
pimaricin biosynthetic enzymes will most probably ensure
that their hybrids are functional. On the other hand,
different specificities of the heterologous modification
35   enzymes might provide new tools for further structural
changes on the molecules produced by genetically
engineered strains.

- 92 -

Example 6: Further manipulations of the regulatory genes leading to increased production of nystatin

Expression of *nysRII* under control of the PermE* promoter
5   in *S. noursei* ATCC 11455

To confirm the function of the *nysRII* gene
predicted through analysis of its coding sequence (see
Example 1), it was expressed in *S. noursei* ATCC 11455
under control of the PermE* promoter. The 2168 bp *SalI*-
10   *BclI* fragment from the phage N58 (representing C-terminal
part of *nysRII*) was cloned into the *SalI*-*BamHI* digested
pGEM11Zf(-), resulting in construct pC1A1. The 811 bp N-
terminal part of the *nysRII* gene was PCR-amplified from
the phage N58 template with the oligonucleotide primers
15   NSR2.1 (5'-GCCGGCATGCGACGAA CAG GACGAGAGGT-3') (SEQ ID
NO. 44.) and NSR2.3 (5'-GCCGTGGTCGACGAA GGC-3') (SEQ ID
NO. 45). The conditions for PCR were as described above.
The PCR fragment was digested with *SphI* and *SalI*, and
ligated, together with the 2168 bp *SalI*-*HindIII* fragment
20   from pC1A1 into *SphI*-*HindIII* digested pGEM3Zf(-) vector,
giving the plasmid pC2A1. From the latter, the 3.0 kb
SphI-*HindIII* fragment was isolated and ligated, together
with the 0.3 kb *EcoRI*-*SphI* fragment containing the PermE*
promoter either with *EcoRI*-*HindIII* digested pSOK804
25   vector (generating plasmid pC3A1), or with the 3.0 kb
EcoRI-*HindIII* fragment from pSOK201 (generating plasmid
pC3E1). Since the pSOK804-based vectors integrate site-
specifically in the *S. noursei* chromosome, the pC3A1
plasmid could be regarded as a construct for *nysRII*
30   expression in-*trans*. Plasmid pC3E1, on the other hand,
is a suicide vector capable of integrating into the *S.
noursei* genome only via homologous recombination through
the cloned *nysRII* gene, thus providing expression of the
latter in-*cis*. Introduction of plasmids pC3A1 and pC3E1
35   resulted in recombinant strains C3A1 and C3E1,
respectively. In the former strain PermE* promoter is
placed upstream of both *nysRII* and *nysRIII* genes (in-

- 93 -

cis), while in the latter strain *PermE** promoter is
placed upstream of *nysRII* gene (in-*trans*). Nystatin
production by the C3A1 and C3E1 strains was increased by
18% and 22%, respectively, compared to the wild-type *S.*
5 *noursei*. Moreover, during fermentation experiments it
was noticed that nystatin production by the C3E1 strain
reached its maximum 24 h earlier than the wild-type
strain. These data support the assumption that the
nysRII gene encodes a positive regulator that may be used
10 for enhancing the production of nystatin and its
derivatives in *S. noursei* strains.

Expression of *nysRIV* under control of the *PermE promoter
in *S. noursei* ATCC 11455:**
15   The start codon for *nysRIV* was reassigned, and is
likely to be located 48 nt upstream of the originally
proposed start nucleotide. Thus, *nysRIV* presumably
encodes a 226 aa (long) rather than a 210 aa (short)
protein as was previously suggested. The long and short
20 versions of the *nysRIV* gene were PCR-amplified from the
N58 recombinant phage DNA with oligonucleotide primers
NR4P3 (5 - CTCAGCATGCCGAAAGGATGGCGG-3') (SEQ ID NO. 46)
and NR4P5 (5'-AGGCAAGCTTCGGCGACACGGGCGT-3') (SEQ ID NO.
47), or NR4P4 (5'-CTCAGCATGCGTACGACCGGCGGG-3') (SEQ ID
25 NO. 48) and NR4P5, respectively. The conditions for PCR
were as described above. The corresponding PCR products
of 0.78 kb and 0.73 kb were digested with *SphI* and
HindIII, and ligated, together with the 0.3 kb *EcoRI-SphI*
fragment containing the *PermE** promoter, with the *EcoRI-*
30 *HindIII* digested pSOK804 vector, resulting in plasmids
pNR4EL and pNR4ES, respectively. Both plasmids were
introduced into *S. noursei* ATCC 11455 generating mutant
strains NR4EL and NR4ES, respectively. Nystatin
production by the strain NR4ES (expressing a 210 aa
35 protein from the *PermE** promoter) did not differ
significantly from that of the wild-type *S. noursei*
harboring only pSOK804, while the NR4EL recombinant

cis), while in the latter strain *PermE** promoter is placed upstream of *nysRII* gene (in-*trans*). Nystatin production by the C3A1 and C3E1 strains was increased by 18% and 22%, respectively, compared to the wild-type *S.*

5   *noursei*. Moreover, during fermentation experiments it was noticed that nystatin production by the C3E1 strain reached its maximum 24 h earlier than the wild-type strain. These data support the assumption that the *nysRII* gene encodes a positive regulator that may be used

10   for enhancing the production of nystatin and its derivatives in *S. noursei* strains.

Expression of *nysRIV* under control of the *PermE** promoter in *S. noursei* ATCC 11455:

15   The start codon for *nysRIV* was reassigned, and is likely to be located 48 nt upstream of the originally proposed start nucleotide. Thus, *nysRIV* presumably encodes a 226 aa (long) rather than a 210 aa (short) protein as was previously suggested. The long and short

20   versions of the *nysRIV* gene were PCR-amplified from the N58 recombinant phage DNA with oligonucleotide primers NR4P3 (5 - CTCAGCATGCCGAAAGGATGGCGG-3') (SEQ ID NO. 46) and NR4P5 (5'-AGGCAAGCTTCGGCGACACGGGCGT-3') (SEQ ID NO. 47), or NR4P4 (5'-CTCAGCATGCGTACGACCGGCGGG-3') (SEQ ID

25   NO. 48) and NR4P5, respectively. The conditions for PCR were as described above. The corresponding PCR products of 0.78 kb and 0.73 kb were digested with *SphI* and *HindIII*, and ligated, together with the 0.3 kb *EcoRI-SphI* fragment containing the *PermE** promoter, with the *EcoRI-*

30   *HindIII* digested pSOK804 vector, resulting in plasmids pNR4EL and pNR4ES, respectively. Both plasmids were introduced into *S. noursei* ATCC 11455 generating mutant strains NR4EL and NR4ES, respectively. Nystatin production by the strain NR4ES (expressing a 210 aa

35   protein from the *PermE** promoter) did not differ significantly from that of the wild-type *S. noursei* harboring only pSOK804, while the NR4EL recombinant

- 94 -

(expressing a 226 aa protein from the PermE* promoter) produced nystatin at a level 36 % above the wild-type level. These data suggest that the longer, 226 aa - long version, represents the actual NysRIV polypeptide.

5   Moreover, these data support the assumption that nysRIV gene encodes a positive regulator that may be used for enhancing the production of nystatin and its derivatives in S. noursei strains.

- 95 -

Claims

1.   A nucleic acid molecule comprising:

(a)   a nucleotide sequence as shown in SEQ ID No. 35;
or

(b)   a nucleotide sequence which is the complement of
SEQ ID No. 35; or

(c)   a nucleotide sequence which is degenerate with SEQ
ID No. 35; or

(d)   a nucleotide sequence hybridising under conditions
of high stringency to SEQ ID No. 35, to the complement
of SEQ ID No. 35, or to a hybridisation probe derived
from SEQ ID No. 35 or the complement thereof; or

(e)   a nucleotide sequence having at least 80% sequence
identity with SEQ ID No. 35; or

(f)   a nucleotide sequence having at least 65% sequence
identity with SEQ ID No. 35 wherein said sequence
preferably encodes or is complementary to a sequence
encoding a nystatin PKS enzyme or a part thereof.

2.   A nucleic acid molecule as claimed in claim 1,
said molecule comprising:

(a)   a nucleotide sequence as shown in SEQ ID No. 1
and/or in SEQ ID No. 2; or

(b)   a nucleotide sequence which is the complement of
SEQ ID No. 1 and/or SEQ ID No. 2; or

(c)   a nucleotide sequence which is degenerate with SEQ
ID No. 1 and/or SEQ ID No. 2; or

(d)   a nucleotide sequence hybridising under conditions
of high stringency to SEQ ID No. 1 and/or SEQ ID No. 2,
to the complement of SEQ ID No. 1 and/or SEQ ID No. 2,
or to a hybridisation probe derived from SEQ ID Nos. 1
and/or 2 or the complements thereof; or

(e)   a nucleotide sequence having at least 65% sequence
identity with SEQ ID No. 1 and/or SEQ ID No. 2, wherein
said sequence preferably encodes or is complementary to
a sequence encoding a nystatin PKS enzyme or a part

- 96 -

thereof.

3.   A nucleic acid molecule as claimed in claim 1 or
claim 2 which comprises a nucleotide sequence having at
5   least 70% sequence identity with SEQ ID No. 35, or SEQ
ID No. 1 and/or SEQ ID No. 2, wherein said sequence
preferably encodes or is complementary to a sequence
encoding a nystatin PKS enzyme or a part thereof.

10   4.   A nucleic acid molecule comprising a part of a
nucleotide sequence as defined in claim 1 or claim 2,
wherein said part is at least 15 nucleotides in length.

5.   A nucleic acid molecule as defined claimed in
15   claim 1, which does not comprise the nucleotide
sequence encoding ORF 1 as set out in Table 1.

6.   A nucleic acid molecule as defined claimed in
claim 1, which does not comprise the nucleotide
20   sequence encoding ORF 2 as set out in Table 1.

7.   A nucleic acid molecule as claimed in any one of
claims 1 to 6 which encodes one or more polypeptides,
or comprises one or more genetic elements, having
25   functional activity in the synthesis of a macrolide
antibiotic or a polyketide moiety.

8.   A nucleic acid molecule as claimed in claim 7,
wherein said macrolide antibiotic or polyketide moiety
30   is nystatin or a nystatin derivative.

9.   A nucleic acid molecule as claimed in any one of
claims 1 to 6, said molecule comprising one or more
genes, and/or one or more regulatory sequences, and/or
35   one or more modules, or enzymatic domains, or non-
coding or coding functional genetic elements of a PKS
gene cluster.

10.   A nucleic acid molecule as claimed in any one of
claims 3 to 9, comprising a nucleotide sequence defined
by any one or more of the nucleotide "start" and "end"
positions within SEQ ID Nos. 35, 1 or 2, as set out in
5   Table 1.

11.   A nucleic acid molecule as claimed in any one of
claims 3 to 9, comprising a nucleotide sequence defined
as lying between any one or more of the nucleotide
10   sequences having the "start" and "end" positions within
SEQ ID Nos. 35, 1 or 2, as set out in Table 1.

12.   A nucleic acid molecule comprising a nucleotide
sequence encoding one or more amino acid sequences
15   selected from SEQ ID Nos 3 to 20 or 36 to 43, or a
nucleotide sequence which is complementary thereto or
degenerate therewith or comprising a nucleotide
sequence which encodes one or more amino acid sequences
which exhibit at least 60% sequence identity with any
20   one of SEQ ID Nos. 3 to 20 or 36 to 43.

13.   A nucleic acid molecule as claimed in claim 12
which encodes one or more amino acid sequences which
exhibit at least 85% sequence identity with any one of
25   SEQ ID Nos. 3 to 20 or 36 to 43.

14.   A polypeptide encoded by a nucleic acid molecule
as defined in any one of claims 1 to 13.

30   15.   A polypeptide as claimed in claim 14, comprising:
(a) all or part of an amino acid sequence as shown in
any one or more of SEQ ID Nos. 3 to 20 or 36 to 43; or
(b) all or part of an amino acid sequence which has at
least 60% sequence identity with any one or more of SEQ
35   ID Nos. 3 to 20 or 36 to 43.

- 98 -

16.   A polypeptide as claimed in claim 14, wherein said amino acid sequence at (b) has at least 85% sequence identity with any one or more of SEQ ID Nos. 3 to 20 or 36 to 43.

17.   A polypeptide as claimed in any one of claims 14 to 16 or claim 11 having functional activity in the synthesis of a macrolide antibiotic or polyketide moiety.

18.   A polypeptide as claimed in any one of claims 14 to 17, comprising a functional region of any one of SEQ Nos. 3 to 20 or 36 to 43, said functional region being as defined in Table 2.

19.   An expression vector comprising a nucleic acid molecule as defined in any one of claims 1 to 13.

20.   A host cell or transgenic organism comprising a nucleic acid molecule as defined in any one of claims 1 to 13 or an expression vector as defined in claim 19.

21.   A method for producing a polyketide or macrolide molecule, said method comprising expressing within a host cell a nucleic acid molecule as defined in any one of claims 1 to 13.

22.   Use of a nucleic acid molecule as defined in any one of claims 1 to 13 in the preparation of a modified PKS system, or in the preparation of a modified polyketide molecule.

23.   A method for preparing a modified PKS system, or a modified polyketide molecule, said method comprising modifying a nucleic acid molecule as defined in any one of claims 1 to 13 or introducing a nucleic acid molecule as defined in any one of claims 1 to 13 into a

- 99 -

further PKS-encoding nucleic acid molecule.

24.   A method as claimed in claim 23, being a method of
preparing a nucleic acid molecule which comprises a
5   nucleotide sequence encoding a modified polyketide
synthase enzyme or enzyme system, said method
comprising (a)   using a nucleic acid molecule as
defined in any one of claims 1 to 13 as a scaffold and
modifying one or more portions of the nucleic acid
10   molecule that encode enzymatic or other functional
activities; or
(b)   introducing one or more portions of the nucleotide
sequence that encode enzymatic or other functional
activities into an alternative (i.e. different
15   "second") PKS scaffold.

25.   A method as claimed in claim 23, being a method of
preparing a nucleic acid molecule comprising a
nucleotide sequence encoding a modified nystatin PKS,
20   wherein said modified nystatin PKS is derived from a
nystatin PKS encoded by a nucleic acid molecule as
defined in any one of claims 1 to 13, said nucleic acid
molecule containing first regions which encode
enzymatic or other functional activities and second
25   regions which encode scaffolding amino acid sequences,
said method comprising:
(a)   modifying at least one said first region; or
(b)   incorporating at least one said first region into
a scaffolding-encoding second region from a different
30   PKS-encoding nucleotide sequence.

26.   A method of preparing a modified nystatin PKS as
defined in claim 25, said method comprising expressing
a nucleic acid molecule prepared as defined in claim 25
35   within a host cell under conditions whereby the
modified nystatin PKS is expressed.

- 100 -

27.   A modified PKS system derived from a nucleic acid molecule as defined in any one of claims 1 to 13 by a method as defined in any one of claims 23 to 26.

28.   A modified PKS system as claimed in claim 27, comprising a modified nystatin PKS wherein one or more of the polypeptides of SEQ ID Nos. 3 to 20 or 36 to 43, or one or more of the enzymatic domains in one or more modules of SEQ ID Nos. 5 to 7, 20, or 37 to 39 are modified.

29.   A modified PKS system as claimed in claim 27, comprising a modified nystatin PKS wherein
        (i)   the ER domain in module 5 of Nys C (amino acids 4953-5239 of SEQ ID No. 7) is inactivated; or
        (ii)   a module of Nys C (SEQ ID No. 7) is deleted; or
        (iii) the DH domain in module 8 of Nys C (amino acids 10086-10289 of SEQ ID No. 7) is inactivated; or
        (iv)   the KR domain in module 7 of Nys C (amino acids 8812-9086 of SEQ ID No. 7) is inactivated.

30.   A multiplicity of cell colonies comprising a library of colonies wherein each colony of the library contains a different modified PKS as defined in any one of claims 27 to 29.

31.   A method of producing a library of PKS complexes or a library of polyketide molecules, said method comprising culturing a library of colonies as defined in claim 30.

32.   A library of PKS complexes or a library of polyketide molecules obtained or obtainable by a method as defined in claim 31.

- 101 -

33. A combinatorial library of polyketides wherein the polyketide members of the library are synthesized by a modified PKS system as defined in any one of claims 27 to 29.

34. A host cell or transgenic organism containing (a) a nucleic acid molecule prepared by a method as defined in any one of claims 23 to 25 or (b) a modified PKS system as defined in any one of claims 27 to 29.

35. A polyketide produced or producible by a host cell or transgenic organism as defined in claim 34.

36. An antibiotic derived from a polyketide as defined in claim 35.

FIG. 1

FIG. 2

FIG. 3

FIG. 4

FIG. 5
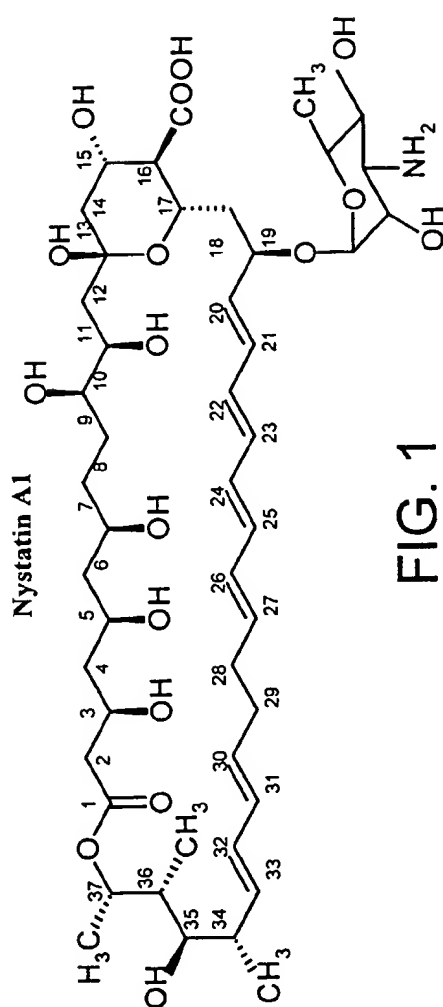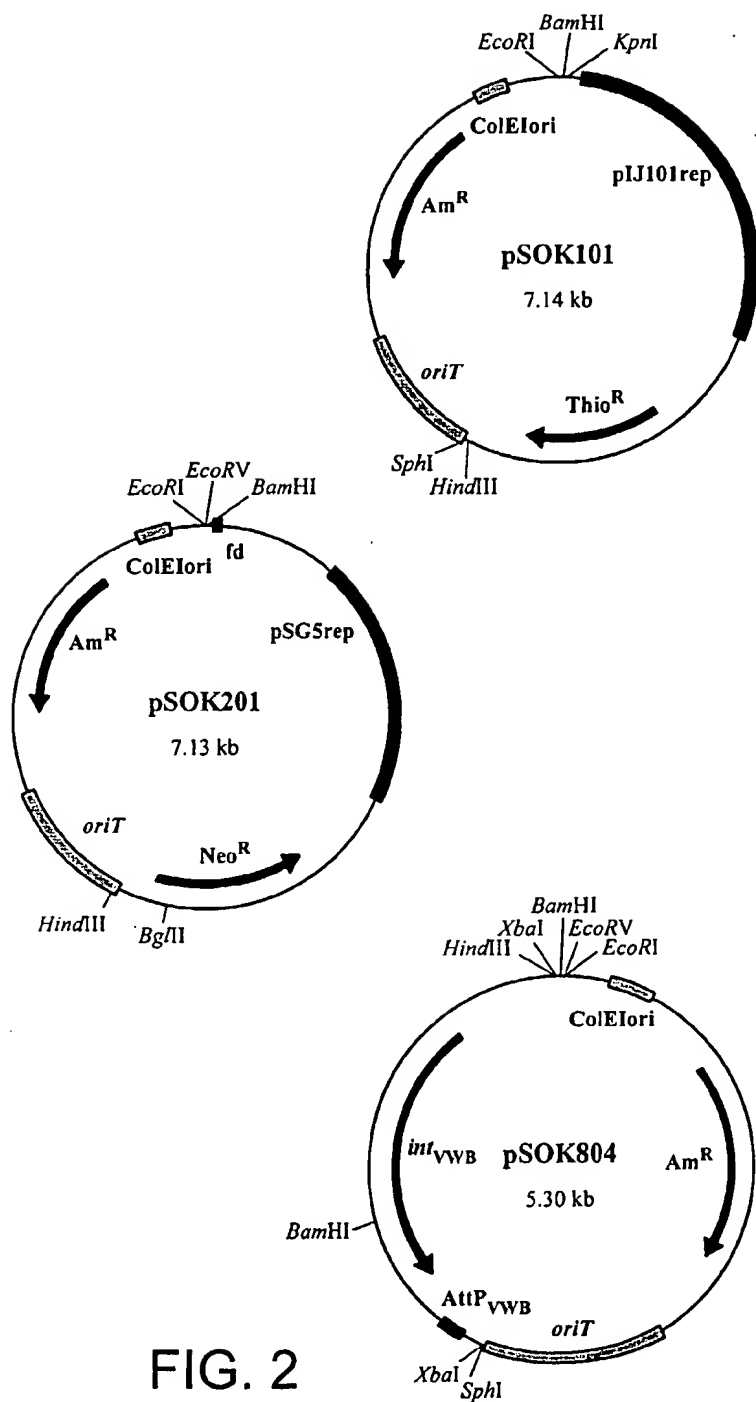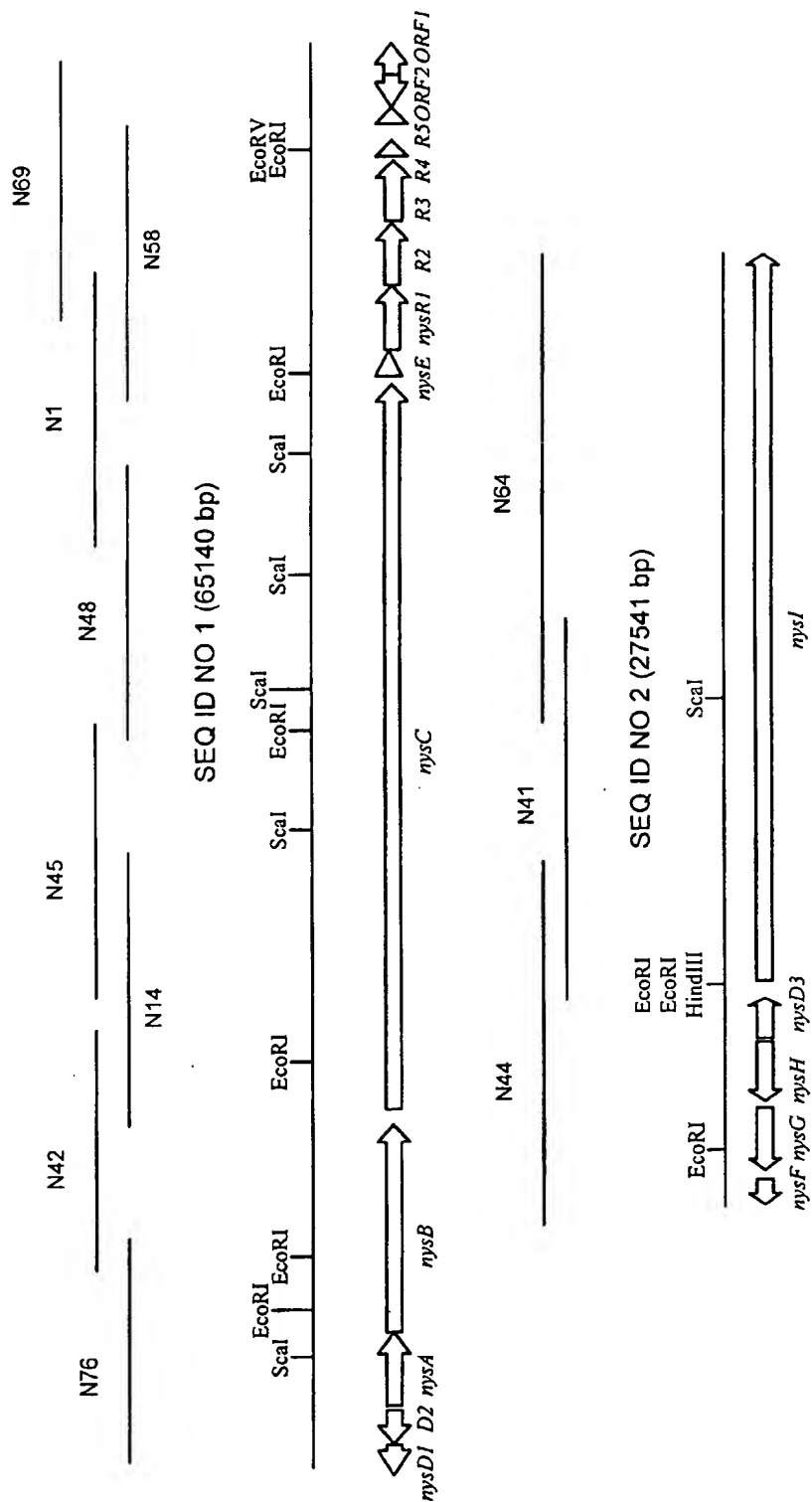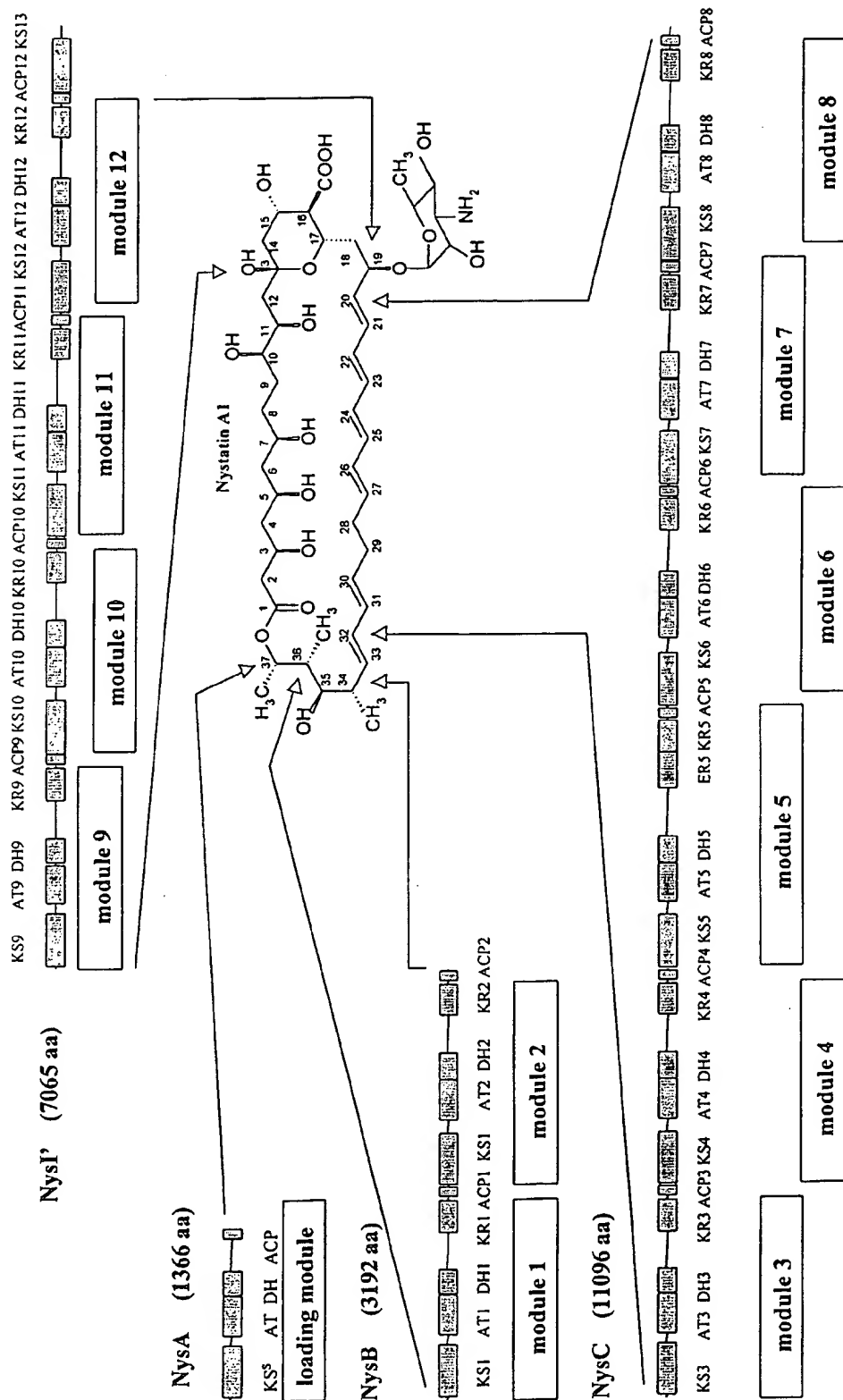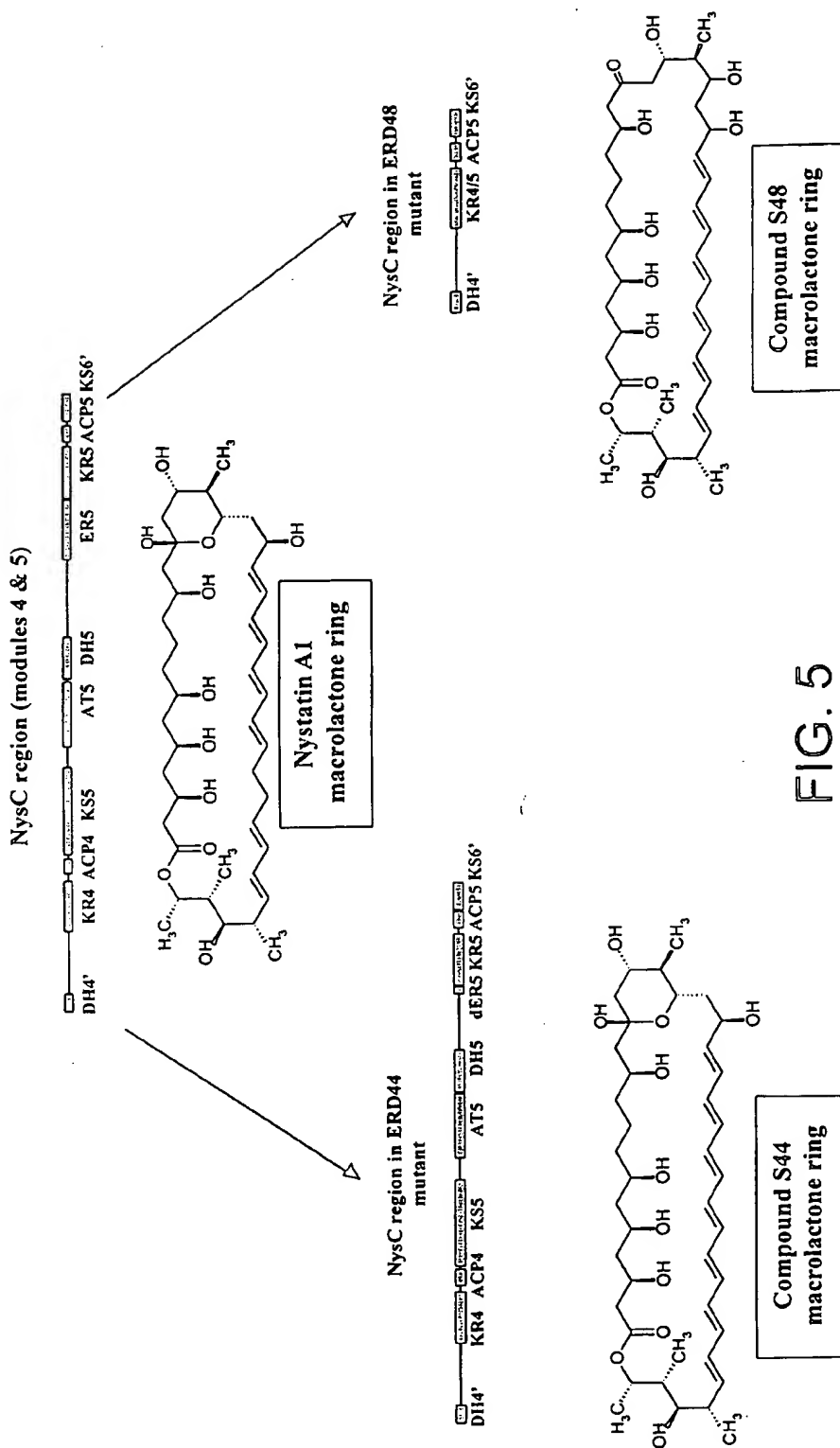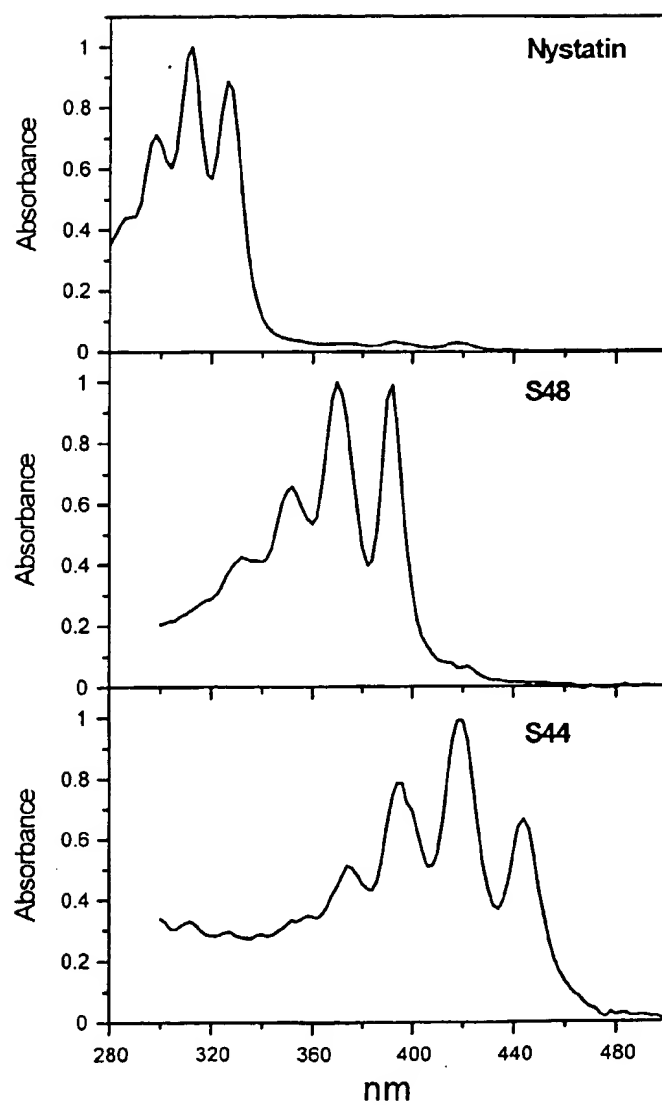
FIG. 6
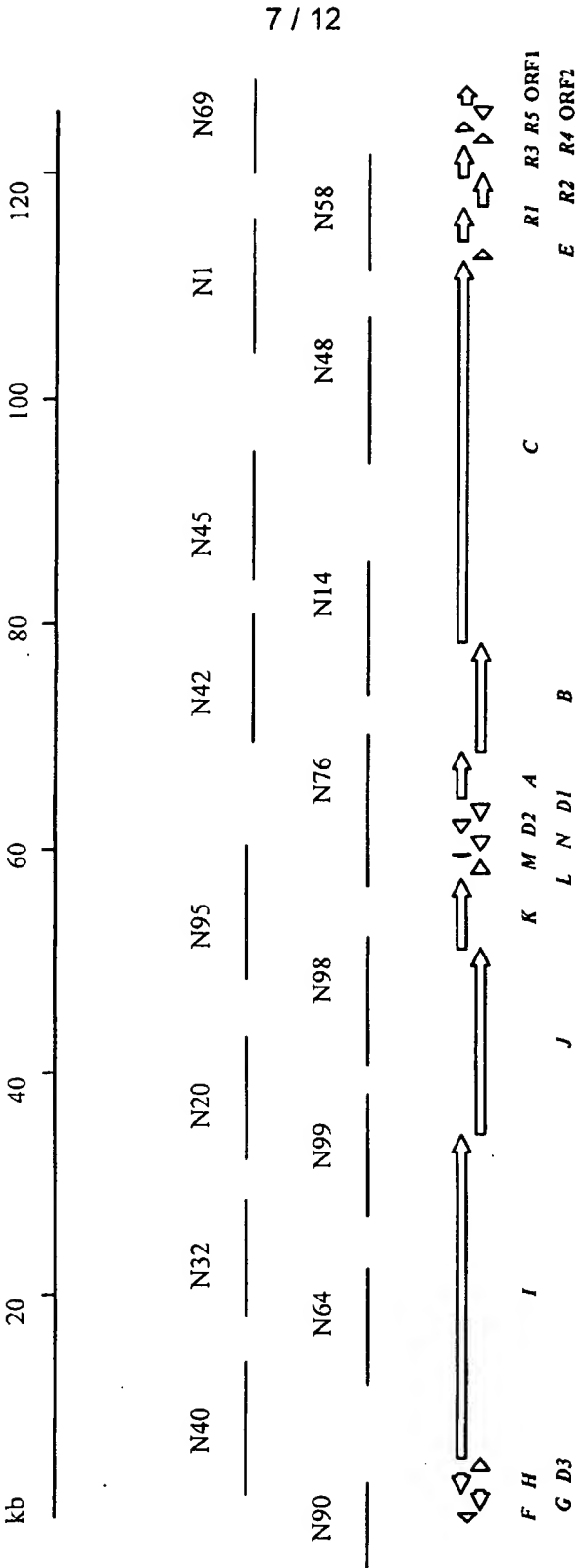
FIG. 7
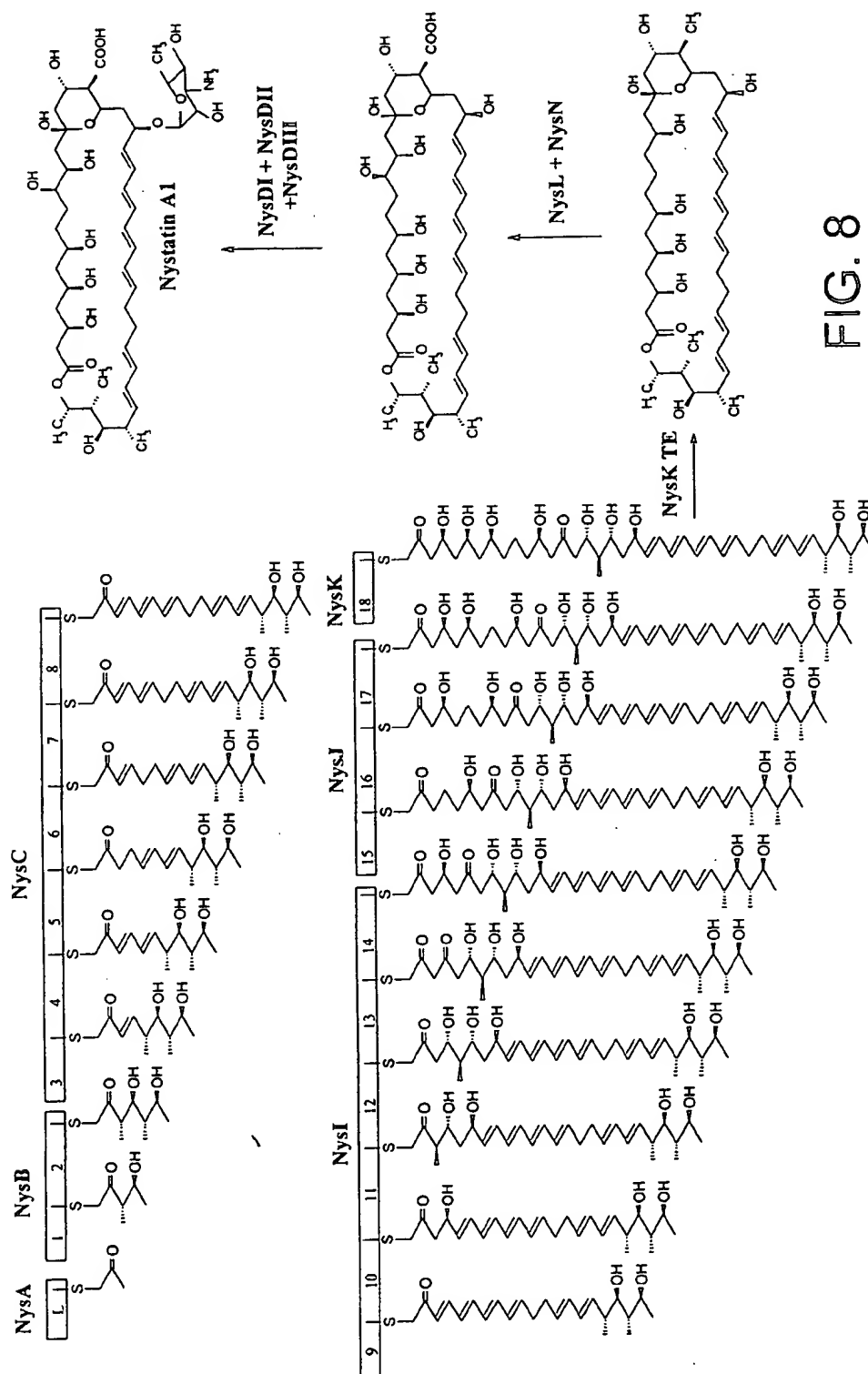
FIG. 8
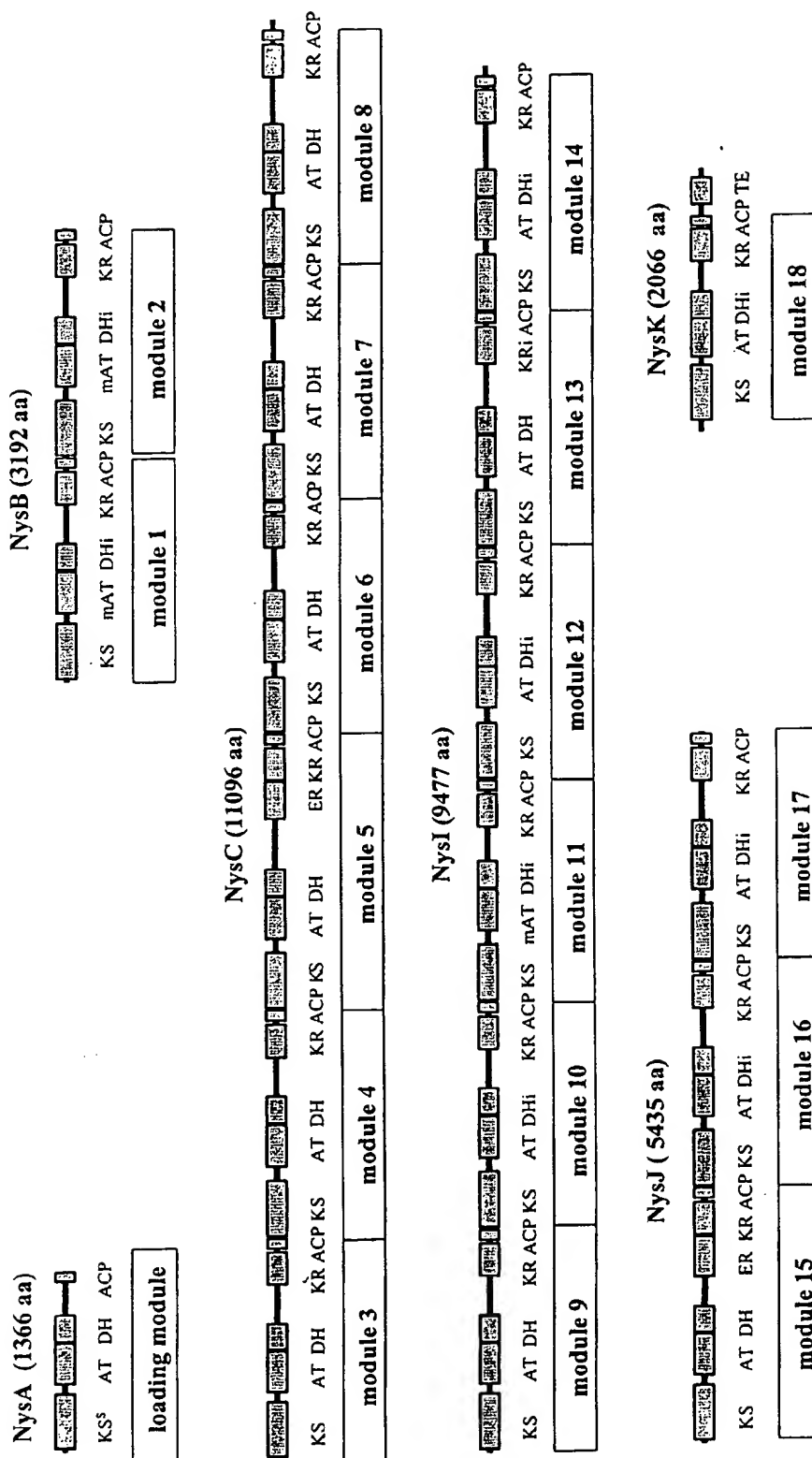
FIG. 9

FIG. 10

11 / 12

FIG. 11

12 / 12
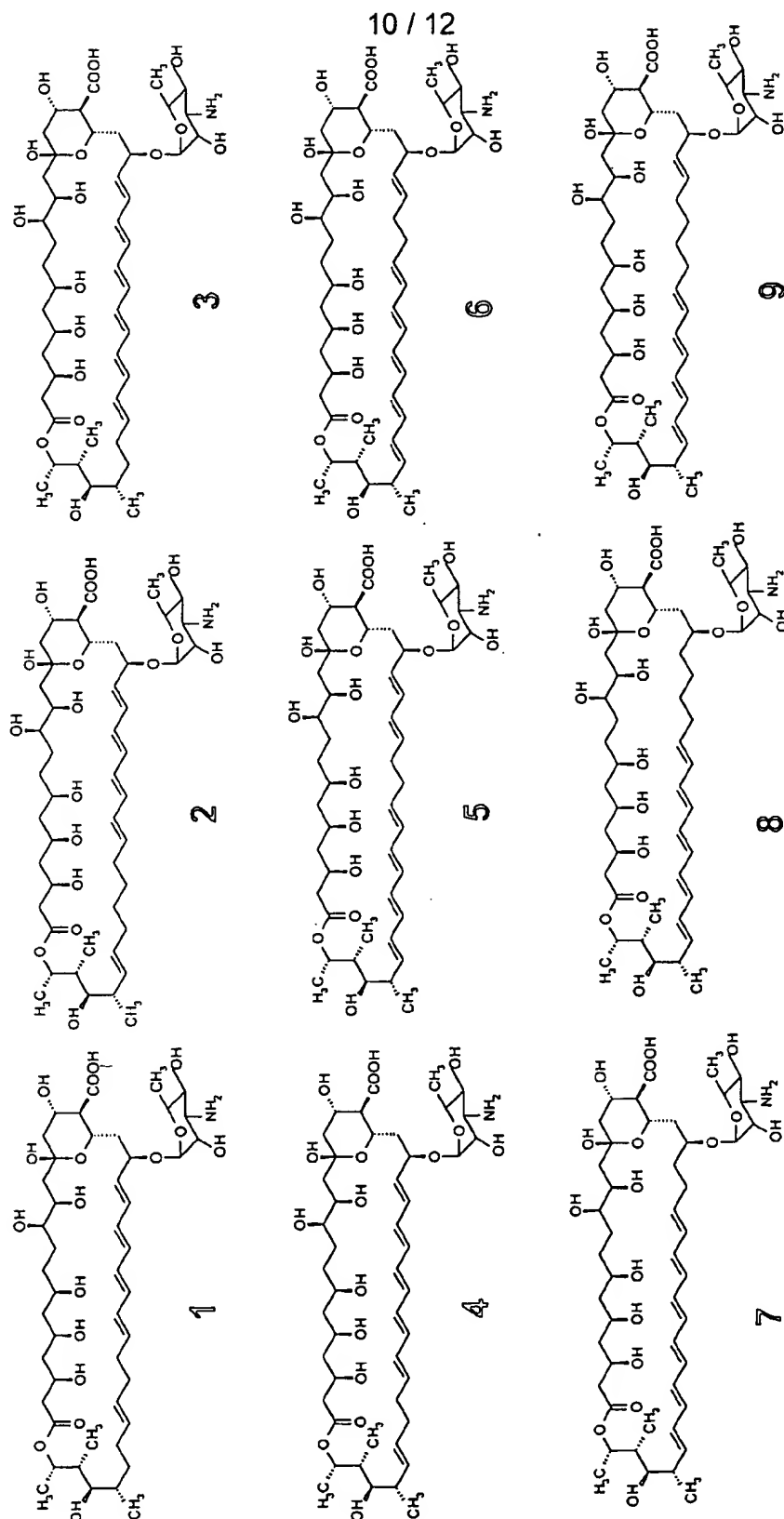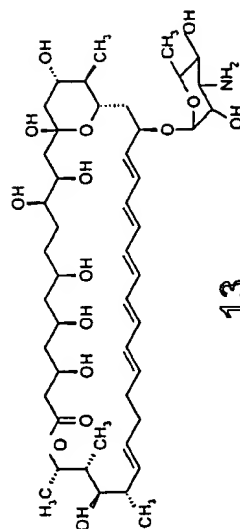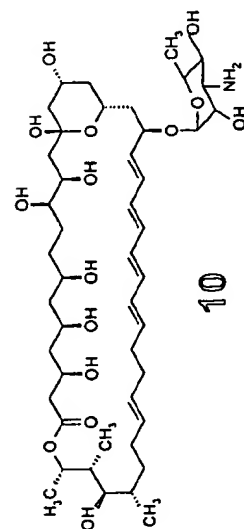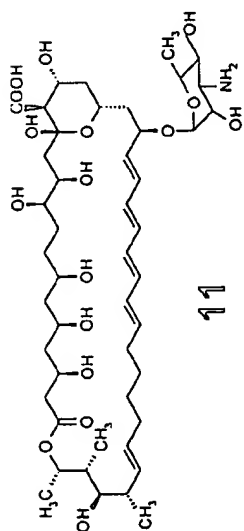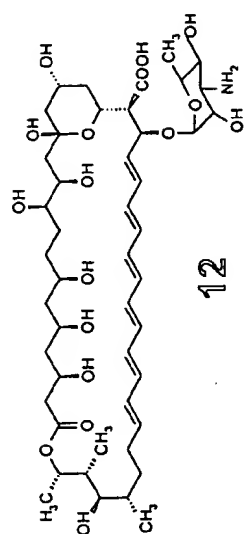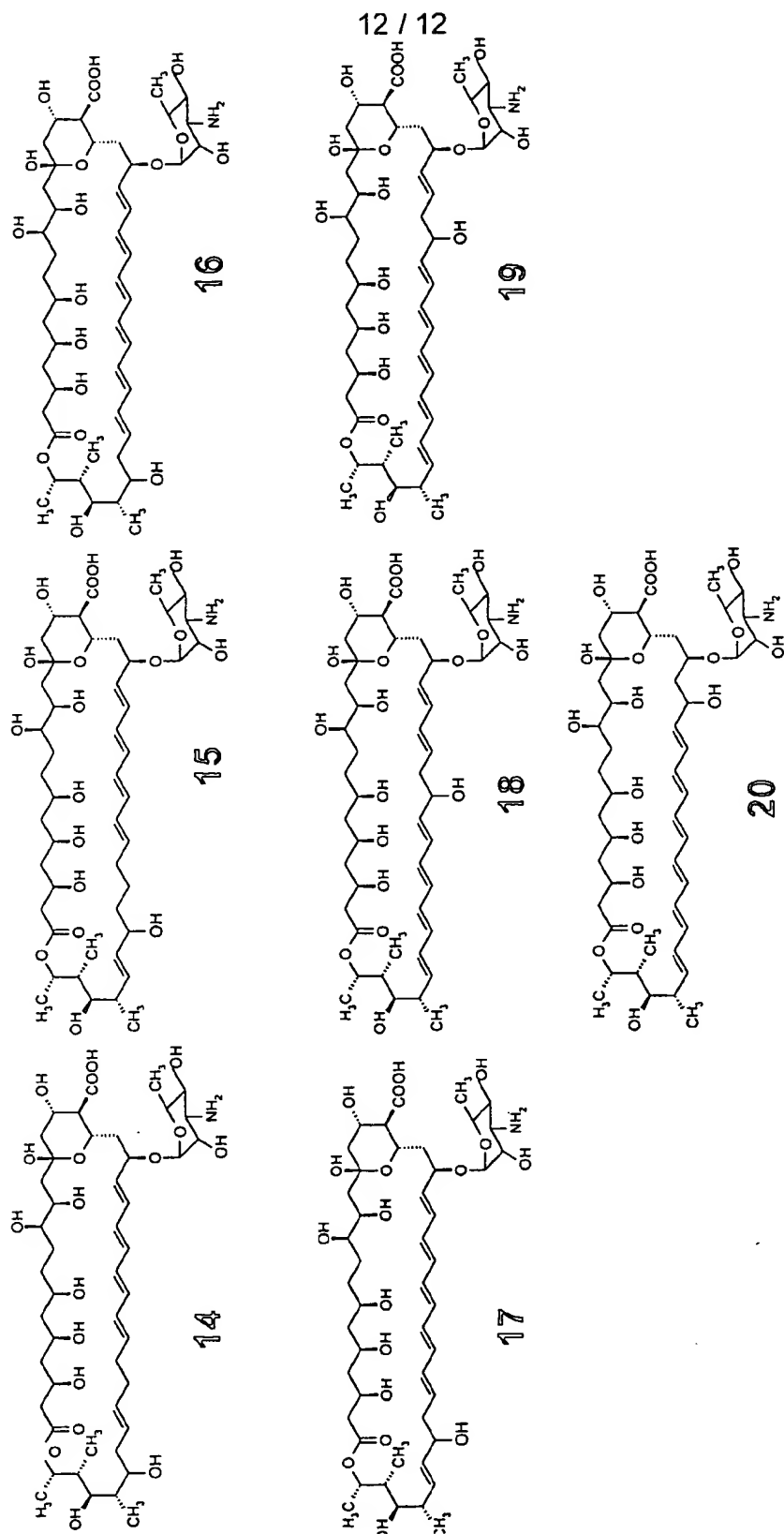
FIG. 12

1

SEQUENCE LISTING

SEQ ID NO. 1

5   DNA sequence: (Nys1; region 1)

```
    1  gatccggccg gtgatgaact cctgctcctg tgcggtcagt ccggtgtgcg
   51  tcggcagata gaagccgtcg gcgctgagcc gggccgcgtt caacgcgggc
  101  cagtcggcgg agaagtaacc gggctgacgg ctcatcggct tgaagaagac
  151  ccgggtctcg atgccctcgc cggcgaggta cgcgcacagc tcgtcccgcc
  201  gctcggcccg caggtcgtac atccacagca cgtcgcgcgg cggcatcagc
  251  gtgatgccgg ggatgtcgcg cagcgcctcg tcgtagcgct tctcgatgtc
  301  gcggcgcagg gcgaggatgg tgtccagctg ctcggtctgc gcgagcgcca
  351  ccgcggcctg catgttggtc atccggaagt tgtaggccag cttcttgtgc
  401  aggaagctgt ggtccttggt gaacgccatc gcccgcagat gggccatctg
  451  ctcggccagg tgcgggtcgt gggtcaggca gacgccgccc tcgccggccg
  501  agatgatctt gttggcgaag agcgagaaac aggcgatgtc gccccgcggc
  551  cgcaccccgt gcgcctcggc ggagtcctcc accacccgca ggttgtactc
  601  gtacgccagg ttcagcacgg cgtccatgtc gcactgccgg ccgtagatgt
  651  gcaccggcat gatcactttg gtgcgcgggg tgatcttctc ctcgatgcgg
  701  gacacgtcga tgttcaggtc gtcgccgcag tccacgaaca ccggcgtggc
  751  accggtgtag gtcaccgccc aggcggacgc gatcatcgtg aactccggga
  801  cgatcacctc gtcaccgggg ccgacgccca gcgcgcgcag cgccagcgtc
  851  agcgccgtgg tgccggagga gcaggcgacg ccgaacggca cgtcgttgta
  901  cgcggcgaac gcctcctcga accgcctgac gtacggcccc tgcgaagaga
  951  tccagccgcc gccgacggcc tccgtcacat agtcgagctc gcggccctgg
 1001  agccacggca tggacaccgg atacgtaaag gacatgggtt ttgagtcttc
 1051  ctcggtcagt cggttgccag ggcgggaagg ccgagcagca ggtccgccgc
 1101  cgcggcccgg ccgccggcgt cccgcagcag cccggcgaag tgctccgcgc
 1151  gctcggtgaa ggagggttgg tcgagcacgc gggtgatctt gtccaggacg
 1201  tcctcggtgt ccacggtctc cggccggtcc agggtcaggc tcaccccgaa
 1251  gtcctggccc cggatcgcct ggtcgtcgca gtccacccac aacggccgga
 1301  ccaccagcgg ctttccgaag tacaggccct cgtggtagcc gttgccaccg
 1351  gcatgggtga agaacgcctt cacgttcgga tgggccagca cgtccagctg
 1401  cgacggcacc cagccctcga tccgcaggtt gtccggcagc tcggcggccg
 1451  gcggcagcaa ctcctgttgg ccgcgcggga gtttccacaa cacctggtgg
 1501  ccccggccgt ccagtcgccg ggcgacctcc accagcgacg ccacctgctc
```

SUBSTITUTE SHEET (RULE 26)

2

```
         1551   acgggtcagc cgggtgatcg tgccgaagcc catgtacacc acggacttct
         1601   gcgccgacag ccagtccgac aggccgtcgt cgtccggtgc ctggggcagc
         1651   ggcggcacca tcgtgcccac cagccgcagc ttcggatgca tcgggaacgg
         1701   gtagtccaac tcccttacgg agtagcacaa gacctgctcc gcatggtcga
    5    1751   tccgcgccat catctgccgc gcctggggcg cgatgcccag ctcggtgcgg
         1801   acccggttgt cctcctcgac gaccttgcgg acgtccgacg tcaggaacat
         1851   cccgagcgtc cgcagccgga acagctggtt ctcgatccgc tgagccaggg
         1901   acatcgcggc cggcagcccc gagtgcggca ccgggaaacc cgacggggtg
         1951   taggacttgg cgaacgggac gtgcgaggtg aggacgttgc tcggcacgaa
    10   2001   cggcaccccg agcacgaacg gaatgccctt ggtgatcgcc agctcgtacc
         2051   cgaactggca catgctctcg atcaccatca gcgccggctc gacctcctcg
         2101   acgatctcct ccaggcggcg gtacttcgcc atccgcgact ccggcgcgaa
         2151   cgaatgccga atcaccgccg cgtgcgcctt gaaccgcgac cgctgcgtca
         2201   cctccgcata cgtcgcgtcg tcccatgtga ccgccgacat ctgcgagacg
    15   2251   gtgtcgccga gcgacgcgaa ccgaaccggg ctgccgtcca ccacggccgc
         2301   cacctcgtcg cgcgctttct cgtcggtggc gaaccacagg tccgccacgt
         2351   cgcgccggga caattccccg gccagcacga gcagcggatt gagcaggccg
         2401   ctttcggcat aactgacgaa caggatcggc cgccgattcg cgcccatgga
         2451   caacacccct cggaatgtgg cgggccgccg ggcccgcgcg ccacgcaccc
    20   2501   gcccggcccg gtcgccgggt gagtgcattc gccgacgccg ccacccgagg
         2551   cgcgtgttgc cggaaggaag ggtcaccggc cggcacccgg aacgcgccgc
         2601   gtggaaaacg ggtcggttac ttggtctcat gccacggacc ggggaatcac
         2651   tagtcttcgg cgcgcgacgg ccctttccgg gccgtgtggc caatgcccgt
         2701   ccccggcgcc cgtcattcct tagggaaaag tacagcgttt gcgaacgtac
    25   2751   gatccggcac gcagaggtga cctgaggcca acttttccgc aggggtgagc
         2801   aaggcatgac gatcggagcc gacgaggacc cggtggtggt cgtcggaatg
         2851   gcctgccgtt atccgggtgg ggtcgccggc ccggaggacc tgtgggaact
         2901   ggtccgcacc ggccgcgacg cgaccaccgc cttcccggac gaccgcggct
         2951   gggacctggc cgcactggcc ggcgacggac ccggccgcag cgcgacccgc
    30   3001   gagggcggat tcctcaccgg cgccgccgac ttcgacgccg ccttcttcgg
         3051   catgtcgccc cgcgaggccg tctccaccga cccgcaacag cgcctcgtcc
         3101   tggagaccgc ctggaagccc ctggagcgcg ccggcatcga cccgcactcc
         3151   ctgcgcggca gccgcaccgg ggtcttcgtc ggcgccagcg gccaggacta
         3201   cgccgccgtc acccacgcct cgcccgacga cctggacgga cacgccctca
    35   3251   ccggcctggc ccccggcgtc gcctccggtc gcctggcgta cgtcctgggc
         3301   ctcgaaggcc ccgccgtcac cgtcgacacc acgtcctcct cgtcgctggt
         3351   cgcgctgcac tgggcggtcc gcgccctgcg cgcggggggag tgcagcaccg
```

3

```
       3401  ccctggccgg cggcgtcacg gtgatgtcca ccccggccgc cttcgtcggc
       3451  cacacccgac agggcggcct cgcgcccgac ggccgctgca agccgttctc
       3501  cgacgacgcc gacggcaccg cctgggcgga gggcgtcggc atcgtcgtcc
       3551  tggagcacct gtccaccgcc cgcgccgccg gcaaccccgt cctcgccgtg
    5  3601  ctgcgcggct cggccgtcaa ccaggacggc gcctccgacg gcctcaccgc
       3651  acccagcggt cccgcccagg aacgcgtcat ccgcgccgcc ctcgccgacg
       3701  cccgactcgc ccccgccgac atcgatctcg tcgaggcgca cggcaccggc
       3751  acccggctcg gcgacccegt cgaggcccgg gcgctgctcg ccgcctacgg
       3801  ccaggaccgg gaccgggacc gaccgctgcg cctcggttcc ctgaagtcca
   10. 3851  ccctcggcca cgcacaggcc gccgccggca tcggcggagt gatcaagacc
       3901  gtcctgaccc tgcggcacgg cctgatgccg cgcatccggc acctggccac
       3951  ccccacccgc caagtcgact ggtcccaggg cgccgtggcc cccctcaccg
       4001  accacacgcc ctggccaccg gccgaccgac cgcgccgcgc cggcgtctcc
       4051  tccttcggca tcagcggcac caacgcccat gtgatcctcg aagaggcgcc
   15  4101  gcccgccgac gtccccgtca cccggcccgg caccctccgc cccagcaccg
       4151  tccctggcc ggtctccgcc gccacgcccg aagccctcga cgcccaactc
       4201  gcccggctcc gcgcccacct gcgcacccac tcggacctgg acccgctgga
       4251  cgtcggctac tccctggcca ccggccgcgc cgcgctccgc caccgggcgg
       4301  tcctcctgcc gcccgccgac ggcaccgccg cggacgccgt cgagcacgcc
   20  4351  cgcggtgcgg cccaccagcg ccgcaccgcc gtcctcttct ccggccaggg
       4401  cagccagcgc ccgggcatgg ccgcgaact cgccgcccgc ttcccggtgt
       4451  tcgccgacgc actggacgac gcgctgcgcg ccctggaccg gcacctggac
       4501  ggcccggtgc gcgaggtgat gtggggcacc gacgccgcgc tcctggaccg
       4551  gaccggctgg acccagcccg ccctgttcgc cgtcgaggtc gccctccacc
   25  4601  gcctggtcgc gtccctcggc gtcaccccccg acttcgtcgg cggccactcc
       4651  gtcggcgaga tcgccgccgc ccacgtcgcc ggcgtcctga gcctggagga
       4701  cgcctgccgc ctggtggccg cccgcgccac gctgatgcag gcgctcccgg
       4751  ccggcggcgc gatggccgcg ctggaggcca ccgaggacga agtggccccg
       4801  ctgctcggcg cacacctcgc gctggccgcc gtcaacggcc ccaccgcggt
   30  4851  cgtcgtcgcc ggagccgagg acgccgtgcg gcaactgacc gcccgcttcg
       4901  ccgaccgcgg ccggcgcacc agccggctgg ccgtctcgca cgccttccac
       4951  tcgccgctga tggagcccat gctcgacgcc ttcgggacg tcgtgagccg
       5001  actgaccttc caccagccgt cgatcccgct ggtctccaac ctcaccggtg
       5051  aactcgccgg cagtgagatc accagcgccg agtactgggt ccggcacgtc
   35  5101  cgcgacaccg tccgcttcgc cgacggcatc accgcactgg ccaaggccgg
       5151  cgccgacgtc ctgatcgaac tcggcccccgg cggcgtgctg tccgcgatgg
       5201  cccgcgacac cctcggcccc gacagcacca ccgacgtcgt ccccgccctg
```

4

```
5251  agcaagggac ggcccgagga gaccgccttc gccggcgccc tcggccgcct
5301  gcacaccctc ggcgtccccg tcgactggcc cgccttctac gccggcaccg
5351  gcgcccgccg cgtcgaactg cccacctacg ccttccagca cgtgcgccac
5401  tggcccaccc cgccccgccc gaacggcgcc gggcccggcg ccctcggcca
5451  cccctgctc  ggctccgccg tcgaactcgc cgacggcggc ggcaccgtct
5501  gctccggcgc cctctccctc cgcacccacc cctggctcgc cgaccacacc
5551  gtcgccgggc gggtcgtgct gccggccacc gcgctgctgg aactcgccgt
5601  gcgcgccggc gacgaggcgg gctgcgacgt cctgcacgaa ctccacctca
5651  ccaccccgcc ggccctgccc gacgacgccg ccctgcacgt ccaggtgcac
5701  gtcggccccg ccgacaccac cgggcgccgc gccgtcaccg tccacacccg
5751  ccccgaccac cacccggccg gcgactggac ccgatgcgcc accggcaccc
5801  tcggcagcac cccgccgtcc gcagccgaag ccgccacggg cggcaccccg
5851  gccgcctggc cgccggccga cgccgaaccc ctcgacctcg ccgaccacta
5901  cgaacggctc gccgaccgcg gcttcgacta cggcccgacc ttccgcggcc
5951  tgcgggccgc ctggcgacgc ggcgcggaga tcttcgcgga cgtggaatgc
6001  ccgcccggca ccgccgacga cgcccccgac cacggactgc accccgccct
6051  gctcgacgcg gcccggcacg ccgccatggc ggtggacggc accgtgcccg
6101  tcgcctggca cggcgtccgg ctgcacgccg tcggcgccac cgcgctgcgg
6151  gtccgcatcc gccccaccac gaccggcacg ctgaccctca ccgcggtcga
6201  cgtgcacggc gcgccggtcg tcaccgtcga ggccctcacc gcccgcccgc
6251  tgaccgacga ggaacgcgcc gccccgcgga cgccgcggca ggcccgcggc
6301  gagacgcccg ccgacgcccg cccggcccgg cccgcggcgg cccgccccgg
6351  cccggccggc gaacccctcc cggacaccac cgggtcccac cccaccgccg
6401  gccacctcgc cgcgctgccg ccggccgccc gggagcgcca gctgctggac
6451  ctggtgcgca cccaggccgc cgccgtcctg ggccacccg  gccccgaggc
6501  cgtcggcacc cgcagcgtct tcaaggagct gggcttcgac tcgttggccg
6551  gcgtcgaact cgccgaccgg ctcaccgccc gcaccggact gcgcctgccg
6601  gccaccctcg tcttcaactt ccccaccccc gaacgtgctg cccaccgcct
6651  cggggaactc ctcgccgcaa ccgcccccct cgaccccggg gcgtacggag
6701  aggaactcac caggttcgag gcgatcgtga cgaacctgcc gcaggacggc
6751  cccgaacgcc gggccgtcgc ggaccggttg gacgccatcg tctccgcact
6801  ccgccagaac tcgcctgcag aggtgccctc ctcggacgag gacatcgaca
6851  cggtgtcggt cgacagactg ctcgacatca tcgatgaaga gttcgaaacc
6901  acatagagaa attgttgctt tcgttcgcga cccgatgacg aggacggacc
6951  gatgcaggaa ccccagcaag gccagccgga ccagcaggag aaaatcgtcg
7001  actatctccg gcgggtcact tcagatcttc gccgtgcccg ccgccgcatt
7051  ggcgaactgg aatccaagga caacgagccc atcgccatcg tcggaatggg
```

5

```
      7101  ctgccgactt cccggcggcg tcaattcgcc ggaatccctg tgggacctgg
      7151  tgcgttccgg cggcgacgcc atttccggat tccccgtcga ccgcggctgg
      7201  gacctggaaa ccctcaccgg aaacggcgac ggcagcagcg ccacccacga
      7251  aggcggattc ctctacgacg ccgcggaatt cgacgccgcc ttcttcggca
  5   7301  tctcgccgcg cgaggcgact gctatggacc cccagcagcg cctcctcctc
      7351  gaagtcgcct gggaggcgct ggagcgcgcc ggcatcgccc ccacagccct
      7401  gcgcggcagc cggtccggcg tgttcgtcgg ctcctaccac tggggcgcgc
      7451  cctcggccga cgccgccacc gaactgcacg ccacgccct gaccggcacc
      7501  gccgccagcg tgctgtccgg ccgcctggcc tacaccctcg gcctcgaagg
 10   7551  cccggccgtc accgtcgaca ccgcctgctc ctcctccctc gtcgccctgc
      7601  acctggcggc ccagtccctg cgcgtcggcg aatcctcgct cgccgtgatc
      7651  ggcggcgtca cgatcctcac cgagccgtcc gtcttcgtcg agttcagcgc
      7701  ccagggcggc ctggcaccgg acggccgctg caaggcgttc tccgacgccg
      7751  ccgacggcac cggttgggcc gagggcgtcg gtgtcctcgt cgccgagcgg
 15   7801  ctctccgacg cgcagcgcaa cggccatccg gtcctcgccg tgctgcgcgg
      7851  ctccgccgtc aaccaggacg gcgcctccaa cggcctgacc gcccccaacg
      7901  gcccctccca ggaacgggtc atccagcagg ccctcgcccg gaccggcctg
      7951  accccgccg acatcgacgc cgtcgaggcg cacggcaccg gcacccggct
      8001  cggcgacccc atcgaggccc aggccctgct cgccacctac ggccagggac
 20   8051  acaccccga ccagccgctg tggctcggct ccctgaagtc caacatcggg
      8101  cacacccagg cggccgccgg cgtcgccggt gtcatcaaga tggtcatggc
      8151  gctgcgccac ggccacctgc cgccgaccct gcacgccgac gcgccctcct
      8201  cgcacgtgga ctggtccgcc ggatcggtac gcctgctgac cgagggccag
      8251  cagtggccgg agaccggacg tccgcgccgg gccgcggtgt cctcgttcgg
 25   8301  catcagcggc accaacgcgc acgccctgct ggaacaggca ccccacccg
      8351  cggacaccgc ggacgccggc gacgacgccg cgcccaccga accggccggc
      8401  gcgcccgccg cgctgccctg gatcgtctcc ggacactccc cgcaggcgct
      8451  gcgcgaccag gccgccgccc tggccgccag ggtcgagacc gaccccgcgc
      8501  tccgccccca ggacatcggg cacaccctgc acaccgcccg cgccctgctc
 30   8551  gaacgacgcg ccgtcgtcgt cgcccccgac cgcgccgaac tcctcgcggc
      8601  tacccacgag ttggccgccg gccggtccgc gaacgccgtc gtcgagggcc
      8651  tcgcggacgt cgagggtcgg acggtgttcg tgttccccgg tcagggttcg
      8701  cagtgggtgg ggatgggggc ccaactcctc gatgagtcgg cggtgttcgc
      8751  ggagcggatt gccgagtgtg cggcggcact cgccgagttc accgactggt
 35   8801  cgctggtcga tgtgctgcgg ggtgtggtgg gtgcgccgtc gttggagcgg
      8851  gtcgatgtgg tgcagccggc gtcgttcgcg gtgatggtgt cgttggctgc
      8901  gttgtggggt tcccgtggtg tgttgccgga tgcggtggtg gggcattcgc
```

6

```
         8951  agggtgagat cgctgccgcg gtggtgtcgg gtgcgctgtc gttgcgggac
         9001  ggggcgcggg tggtggcgct gcggagtcag gccattggtc gtgcgttggc
         9051  ggggcggggc gggatgatgt ccgtcgcgct gtcggtggac gtgctcgaac
         9101  cgcggttggt cgagttcgag gggcgggtgt cggtggccgc cgtcaacggc
    5    9151  ccgcgctccg tcgtggtcgc cggcgagccc gaggcgctgg acgcgctgca
         9201  cgcccggctg accgccgacg acatccgggc ccgccggatc gcggtggact
         9251  acgcctcgca ctcgcaccag gtcgaggacc tgcacgagga actgctggag
         9301  gtgctggcgg agctggcgcc gcgcacgtcg gaggtgccgt tcttctcgac
         9351  cgtgaccggc gactggctgg acaccgcgcg gatggacgcc ggctactggt
    10   9401  tccgcaacct gcgcggacgg gtgcggttcg cggacgcggt ggcggacctg
         9451  ctggcggcgg agtaccgcgc gttcgtcgag gtcagctcgc acccggtgct
         9501  gacgatggcg gtcttggacc tgatcgagga ggccgggggtc acggccgtcg
         9551  cgaccggcac cctgcgccgt gaccagggtg gcgcgggccg cttcctgctg
         9601  tcggccgccg aggtcttcgt gcgcggtgtg gacgtggact gggcgggggc
    15   9651  gttcgagggg accggtgcgg cccgggtcga cctgcccacc tacgccttcc
         9701  agcgcgagcg gtactggaac acccgcaccg ccgccgaccg caccccggcc
         9751  gacgccccga tggacgccga attctgggcc gccgtcgaac aggcggacgt
         9801  ctccgcgctg accgccgcgc tcggcaccga cgaggactcc gtcgccgcca
         9851  tcctgcccgg cctcacctcc tggcgccggg cccgctccca gcgcaccacc
    20   9901  ctcgactcct ggcgctaccg cgtcacctgg acgcccctcg cccaggtgcc
         9951  ccgcgccacc ctgaccggca cctggctgct ggtcaccacc gacggcatcg
        10001  acgacaccga tgtggcaggg gcgttggaga gctacggcgc cgaggtgcgc
        10051  cggctggtcc tggacgagga gtgcaccgac cgcgccgtcc tgcgggagcg
        10101  gctggccggc gcggaggacg tgaccggcat cgtctccgtc ctcgccgccg
    25  10151  ccgaggacga cgccgcacgc caccccggcc tcacccgggg actcgcgctc
        10201  accgtctccc tcgtccaggc cctgggcgac gccgaggcga ccgcgccgct
        10251  gtggttcctg acccgcggcg ccttcgccac cggcccgtcc gaccccgtca
        10301  cccggcccct gcagagccag atcgcgggcg tcggctggac caccgcgctg
        10351  gagcacccgc agcgctgggg cggcaccgtg gacctgcccg acaccctcga
    30  10401  cgcccgggcc gcccagcggc tcgccgccgc gctgtccggc gccctcggcg
        10451  ccgaggacca gctcgccgtc cgcgccgccg gggtactggc ccgccgcatc
        10501  gtgcgtgccg acaccgcgc cggacgaccg gcacggacct gggcgccgcg
        10551  cggcaccacc ctgatcaccg gcggctccgg caccctcgcc ccgcagctcg
        10601  cccgctggct ggccgaacgc ggcgccgagc acgtggtgct ggtcagccgg
    35  10651  cgcggtgccg acgcccccgg agcgcccgaa ctcatcgcgg aggcagccga
        10701  gtcgggcacc gaggtgaccg tcgccgcctg cgacatcacc gaccgcgacg
        10751  cggtcgccgc gctgctggcc gacctcacgg ccgacggccg caccctgcgc
```

SUBSTITUTE SHEET (RULE 26)

```
          10801   accgtcatcc  acgccgccgc  cgccatcgag  ctgtccgcgc  tcgccgacac
          10851   caccgtggcg  gagttcgccg  acgtcgtgca  cgccaaggtc  accggcgcac
          10901   ggatcctcga  cgaactgctc  gacgacgcgg  aactggacga  cttcgtcctg
          10951   tactcctcca  ccgccggcat  gtggggcagc  ggcgtgcacg  ccgcctacgt
5         11001   cgccggcaac  gcctatctgt  ccgcgctcgc  cgagcagcgc  cgcgcccgcg
          11051   gactgcgcac  cacctccatc  cactggggca  agtggcccga  cgaccgggca
          11101   cgcgagctgg  ccgacccgca  ccggatccgc  cgcagcggtc  tggagtacct
          11151   cgaccccgag  ctggcgctca  ccgcgctcca  gcacgtcctg  gacgacgacg
          11201   agaccgtcat  cggcctcatg  gacatcgact  gggacaccta  ccacgacgtg
10        11251   ttcaccgcgg  gccggcccgc  gcacctcttc  gaccagatcc  ccgaggtgcg
          11301   gcgccgcctc  gaccaggcat  ccgtcccgga  ccccgcgggc  ccggccgccg
          11351   acggcctggc  cgcccgcctg  cacggcctcg  ccgccgccga  acaggaccgg
          11401   ctgctgctca  ccctggtccg  caccgaggcc  gccgccgtcc  tcggccacgc
          11451   ctcggccgag  tccttccccg  agcgccgcgc  cttccgtgac  ctcggcttcg
15        11501   actcggtcac  cgccgtggac  ctgcgcaacc  ggctcgtggc  cggcaccgga
          11551   ctgcggctgc  cctcgacgat  ggtcttcgac  caccccaact  gcgcggcgct
          11601   cgccgcgttc  ctgaagacga  cggcgctcgg  cgtccccggc  gccgcaccgc
          11651   agcagcacgc  cgctaccggc  accccggccg  acgacgaccc  gatcgccgtg
          11701   atcggcatga  gctgccgcta  ccccggcggc  gccgccaccc  ccgaggaact
20        11751   gctgcggctc  gccctcgacg  gcgccgacgt  catctcggag  ttccccgcgg
          11801   accgcggctg  ggacgcccgg  ggcctgtacg  acccggaccc  cgaccgcccc
          11851   ggccacacct  actccgtcca  gggcggcttc  ctccacgagg  ccgccggctt
          11901   cgatcccggc  ttcttcggga  tctccccgcg  cgaggcggtc  gccatggacc
          11951   cgcagcagcg  gctcctgctg  gagacctcct  gggaggcgtt  cgaacgcgcc
25        12001   ggtatcgacc  ccgcgtcact  gcgcggcagc  gccgccggca  ccttcttcgg
          12051   cgccagctac  caggactact  cctccaccgt  gcagaacggc  acgggggagt
          12101   ccgaggcgca  catggtgacc  ggcaccgcgg  ccagtgtcct  gtccggccgg
          12151   gtctcctacc  tgctcggcct  ggagggcccc  gcggtcaccg  tggacaccgc
          12201   ctgctcctcc  tcactggtcg  ccctgcacct  ggcctgccag  tccctgcgcg
30        12251   acggcgagag  ctccctcgcg  ctggccggcg  gtgcggccgt  gatggccacc
          12301   ccgcacgcgt  tcgtcggctt  cagccggcag  cgtgccctgg  ccaaggacgg
          12351   ccgctgcaag  ccgttctccg  acaccgccga  cggcatgacg  ctcgccgagg
          12401   gcgtcggcgt  cgtcctgctg  gagcgcctgt  cccacgcccg  cgccaacggg
          12451   caccgggtac  tggccgtgat  ccgcggttcc  gccgtcaacc  aggacggcgc
35        12501   ctccaacggc  ctgaccgcgc  ccaacggccc  gtcccagcag  cgcgtcatcc
          12551   gccaggcgct  cgccaacgcc  ggcctgaccg  gcgccgacgt  cgacgcggtc
          12601   gaggcgcacg  gcaccggcac  caagctgggc  gaccccatcg  aggcccaggc
```

8

```
          12651  cctgctcgcc  acctacggcc  aggaccgcga  cgccgaacgg  ccgctgctgc
          12701  tgggctcggt  gaagtccaac  atcggccaca  cccaggccgc  ggccggcgtc
          12751  gccggcgtca  tcaagatggt  gctggccatg  gacgccggcg  aactgcccgg
          12801  caccctgcac  ctcgacgcgc  cctccagcca  cgtcgactgg  accgccggcg
     5    12851  ccgtggaact  gctgcgcggg  cgcaccccgt  ggcccgagag  cgggcgcccc
          12901  cgccggggccg  gtgtctcctc  gttcggcatc  agcggcacca  acgcccacct
          12951  gatcctcgaa  caggccccgg  ccaccgagcc  gccagccgac  cccgaccgcc
          13001  tccgggacac  cgccaccgac  accgtcgtcc  cctggccgct  cgccgccaag
          13051  tccccggccg  ccctgcgcgc  ccaggccgcc  cggctcctcg  ccaccgtcga
    10    13101  gcacgacccc  gacctcccgc  ccgccccgt   gggccacgcc  ctggccacca
          13151  cccgcgccgc  cctcgaacac  cgcgccgtcg  tcgtcggcga  gcgccgcgag
          13201  gacttcctgc  gcggcctggc  cgccctgtcc  accggcgcct  cgacggccgg
          13251  cctggtcagc  ggcatcgccg  gccccgaccc  cgagggagcg  gtcttcgtct
          13301  tccccggcca  gggatcccag  tggtggggaa  tgggccgcga  actcctcgcc
    15    13351  acgtccgagg  tgttccgcac  cgcgatcgat  gactgcgcga  cggccctcgc
          13401  cccgtacgtc  gactggtcgc  tgcacgacgt  cctggccggc  gagggcgacc
          13451  ccgccctgct  ggagcgggtg  gacgtggtcc  agcccgcgct  gttcgccatg
          13501  atggtcgggc  tgtccgcgct  ctggcgctcc  cacggcgtcg  tcccggcggc
          13551  cgtggtcggc  cactcgcagg  gcgagatcgc  cgcggcctgc  gtcgccggag
    20    13601  ccctcagcct  ggccgacgcc  gcccgcgtgg  tggcgctgcg  cagccaggca
          13651  ctgccgcaac  tgtccggacg  cggcggcatg  atgtcggtct  ccgcccccgt
          13701  agagcgggtc  accgcactcc  tcgccccgtg  gcaggaggcg  ctgtccgtcg
          13751  ccgcggtcaa  cggcccctcg  tccgtggtcg  tctccggcga  caccgacgcg
          13801  ctcgacgccc  tgcacaccgc  ctgccaggaa  cagggcgtgc  gggcccgcaa
    25    13851  ggtgtccgtg  gactacgcct  cgcacgggcg  gcacgtcgag  gccgtccgcg
          13901  acgaactcgc  ccgcgtcctc  gcgccggtcg  acccgcgcgc  ccccgaggtg
          13951  ccgttctact  cgacggtcac  cggcgaccgc  gtggacgacg  ccgccttcga
          14001  cggcgcctac  tggtacacca  acctccgcca  gaccgtccgc  atggaggagg
          14051  ccaccccgcgc  cctcctcgcc  gccggacacc  gcgtcttcat  cgaggtcagc
    30    14101  ccgcacccgg  tgctcgccgc  cccgatccag  gagacgcagg  aggccgtagc
          14151  ggaggccacc  ggcgggtccg  cggtggtcct  cggctcgctc  cgccgcgacg
          14201  agggcggccc  gcggcgcttc  ctgacgtcgc  tcgccgaggc  ccacacccac
          14251  ggcgccccgg  tcgactggac  caccaccttc  gcccggtccg  cctaccagcc
          14301  ggtggacctg  ccgacctacc  ccttccaacg  acaggacttc  tggcccgagg
    35    14351  cccggcccgc  caccccggcc  gccggcgccg  acgcgtccga  cgccgcgttc
          14401  tggcaactgg  tcgagaacca  ggacctcgcc  gcgctcgccc  acgcgctcgg
          14451  cgtccccgcc  gacgacgagc  acaccgcgct  cggcaccgtg  ctgccggccc
```

9

```
        14501  tgtccgcctg  gcgcgccaag  gcccaggccc  gcacccggat  cgacgaactc
        14551  cgctaccacg  tccagtggac  ccgggtcgcc  gagcccgcgg  cggcccccac
        14601  caccggccgg  ctgctggtcg  ccgtcccgcc  ggaccacgcc  gacgcccct
        14651  gggtcgccgc  ggcgctcgac  gccctgggca  ccgacaccgt  ccgcttcgag
    5   14701  gccaagggca  ccgaccgcgc  gggatgggcc  gcacagatcg  cccaactcgt
        14751  cgaggacggc  gaggagttca  ccggcgtggt  gtcgctgctg  gccgccgccg
        14801  aggatctcca  cccggacttc  ggctcggtac  cgctgggggct  ggggcagacc
        14851  ctcgtcctcg  tccaggccct  cggcgacgcc  ggcctgaccg  cgccctgtg
        14901  gtgcctgacc  cgcggcgccg  tcgccaccgg  ccgcgacgac  gccctcgaca
   10   14951  gcccgaccca  gggcgccctg  tggggcctcg  gccgggtcgt  ggccctggaa
        15001  caccccgacc  gctggggcgg  cctgatcgac  ctgcccgcca  ccctcgacgc
        15051  ccgcgccgcg  gcccgcctca  ccggcctgct  cgccgacccc  gccggtgagg
        15101  accaactcgc  cgtccgcgcc  accggcgtgc  tcgcccgccg  catggtgcac
        15151  gccgcgccgt  ccgcgccccg  caccgggcgc  cgctggcggg  gccgcggcac
   15   15201  ctgcctgatc  accggcggca  ccggcggcat  cggcggccgg  gtcgcccgct
        15251  ggatggccga  gcacggcgcc  gcccacctgg  tcctgaccag  ccggcgcggc
        15301  ccggacgcac  ccggcgccgc  cgcactccgg  gccgaactgg  aggccctggg
        15351  cgccgcgtc  accctcgccg  cctgcgacgt  cgccgaccgc  gacgccctgg
        15401  ccgccctgct  ggccgacctc  cccgccgacc  agccgctcac  ctccgtcttc
   20   15451  cactccgccg  gcgtggccga  cggggacgcc  cgggcagccg  acctgaccct
        15501  agatcagctc  gacgcgctgc  tgcgcgccaa  actgaccgcc  gcccaccacc
        15551  tgcacgagct  gaccgccccc  ctcgacctcg  acgcgttcgt  gctcttctcc
        15601  tccggcgccg  cggtctgggg  cagcggcggc  cagcccggct  acgccgccgc
        15651  caacgcctac  ctcgacgccc  tcgccgccca  ccgcaggtcc  ctcgacctgc
   25   15701  ccggcgcgtc  cgtcgcctgg  ggcacctggg  gcgaggtcgg  catggccacc
        15751  gtccccgagg  tccacgagcg  actgcaccgc  caagggggtcc  gcgccatgga
        15801  accggaccac  gcgatcggcg  cgctccagca  gatgctggag  gacgacgaca
        15851  ccaccctcgc  cgtgaccctc  atggactggg  aggcgttcgc  accgagcttc
        15901  accgcgaccc  gacccagcgc  cctgttcagc  acggtgcccg  aagccgtccg
   30   15951  cgcggtgacc  ggcgacccgg  gcaccacggc  cggcgacgac  gtggactccg
        16001  cgaccccgcc  gctccgccgc  cacctggagg  agctgtccgc  cgccgagcgc
        16051  ggccgggccc  tggtcgaggc  ggtccgcgcc  gaggcgtccg  cgaccctcgg
        16101  ccacgacacc  cccgacgcca  tccccgccgg  ccgtgccttc  cgcgacgtcg
        16151  gcttcgactc  ggtcaccgcc  gtcgaactgc  gcaaccggct  gcgcaccgcc
   35   16201  ctcggcctgc  cgctgccggc  cgcgctcgtc  ttcgaccacc  ccacccccac
        16251  ggccctcgcc  ggccacctcg  gcgcgctgct  cttcggcacc  gccccgagg
        16301  acgccggcac  cggccgcccc  gacgaccccg  acgcccgcat  ccgcgaggcg
```

```
     16351  ctcgccaccg tccccatcgg acggctgcgc aaggcgggcc tcctcgacat
     16401  ggtgctgaaa ctcgccgacg gagacgcgac cgacgccccc gcccccgagg
     16451  ccgacgcccc ctcggaatcc ctcgacgaca tggacgccga agccctgctg
     16501  cggctggcca ccgagaactc ggcgaactga aagagagctg gagcccacca
  5  16551  tgagcacgaa ccccgacaag tacgtcgagg cactccggtc gtccctgaag
     16601  gagatcgagc ggctgcgccg gcagaacgaa cagctggtgg cggcggcggt
     16651  cgagcccgtc gcggtcgtcg gcatcggctg ccgcttcccc ggcggcgtca
     16701  cctcccccga ggacctgtgg gagctggtcg ccgaggggcg cgacgtcatc
     16751  gggccgttcc cgcaggaccg cggctgggac ctggagaagt tggccggcgg
 10  16801  cggcgagggc ggcagcctcg cgcaggtcgg cggcttcgtc gaggacgccg
     16851  ccggcttcga ccccggcttc ttcggcatct ccccgcgcga agcggtcgcc
     16901  atggacccgc agcagcgcat cctgctggag atcacctggg aagccctgga
     16951  acgcgccggc atcgacccgt ccaccctgcg cggtaccccc accggcgtct
     17001  tcgtcggcac caccggccag gactacggcg aggtcatcaa ggcgtccgcc
 15  17051  gaggacgtcg aggtctactc gaccaccggc cacgccgcca gcgtcatctc
     17101  cggccggctc tcctacaccc tcggcgccga gggcccggcc gtcaccgtcg
     17151  acaccggctg ctcctcgtcc ctggtcgccc tgcactgggc cgtccaggcg
     17201  ctgcgcggcg gcgagtgctc catggccctg gccggcggcg cgtccatcat
     17251  ggccacccccg ggcccgttcg tcgccttcac cgcgcagagc ggcctggccg
 20  17301  ccgacggccg ctgcaagccc ttctccgacc gggccgacgg caccggctgg
     17351  ggcgagggcg ccggcatgct ggtcctgatg cggctctccg acgcccagcg
     17401  cgagggccgc ccggtcctcg ccgtgctgcg cggctccgcc atcaaccagg
     17451  acggcgcctc caacggcctg accgctccca acggcccctc ccagcagcgc
     17501  gtcatccgcg ccgcgctgga cagcgcccac ctcaccgccg ccgacatcga
 25  17551  cgccgtcgag gcccacggca ccggcaccac cctcggcgac ccgatcgagg
     17601  cccaggcgct cctggcgacc tacggacagg accggccgcg gcccctgtgg
     17651  ctcggctcgg tgaagtccaa catcggccac acccaggccg cctccggtgc
     17701  cgccggcgtg atcaaaatga tcatggcgtt gcagcgcggc gtgctgccgc
     17751  gcagcctgca cgccaccgaa cccaccacgg acgtcgactg gaccgccggc
 30  17801  tccgtcgacc tcctcgacga gacggtcgcc tggcccgaga ccggacgcgc
     17851  ccgccgcgcc ggcgtctcct ccttcggcat cagcggcacc aacgcccacg
     17901  tcatcctcga acaggccccc accgccccccg aagagcccac caccgaaccc
     17951  accgtccgcc ccgccgtcgt cccgtgggcg ctctccgccc gcaccgccgc
     18001  cgccctcgac gcccagcgcg cccgcctcac cggccacctc gccgacaccc
 35  18051  ccgacgccga cccccctcgac gtcggctacg cgctcgccga cggacgcgcc
     18101  accttcgaac accgcgccgt cctgctcccc gacggcaccg aactcgccca
     18151  cggaaccgcc ggcgaaggcc cctgcgccgt cctcttctcc ggccagggct
```

```
        18201  cccagcgccc gggcatggga cgcgaactcc acgcccgctt cccggtgttc

        18251  gccgcggcct tcgacgagat cacagcgctc ctcgacaccc acctcgaccg

        18301  cccgctgcgc gaggtcgtct ggggcaccga cgccgacctg ctgaacgaca

        18351  ccggctgggc ccaacccgcc ctgttcgccg tcgaggtcgc cctctaccgc

     5  18401  ctggtcgcgt ccctcggcgt caccccgac ttcgtcggcg gccactccat

        18451  cggcgagctc gccgccgcgc acgtcgccgg ggtcctctcc ctcgaagacg

        18501  cctgcaccct cgtcgccgcc cgcgcccgcc tcatgcaggc cctgccgcgc

        18551  ggcggcgcga tgctcgcgat ccgcgccacc gaggacgagg tcacccccca

        18601  cctcaccgac gacgtctcga tcgccgccgt caacgggccc acctccgtcg

    10  18651  tcgtcgccgg caccgaggaa gccgtcgccg cgatcggggc gcgcttcacc

        18701  gcccaggacc gcaagaccac ccggctgcgg gtcagccacg ccttccactc

        18751  gccgctcatg gacccgatgc tggcggaatt ccgcgccgtc gccgcgggcc

        18801  tgacctacca cgagccgcgc atcccggtcc tctccaacct caccggcacc

        18851  gtcgccgccg tcgccgacct gtgctccgcc gactactggg tccgccacgt

    15  18901  ccgcgaggcg gtccgcttcg ccgacggcgt caccgccctc accgaccgcg

        18951  gcgtgaccac gctcgtcgaa ctcggcccgg acggcgtgct gtccgccatg

        19001  gcccaggaat ccctgccgga cggcgccgcc gccgtgccgc tgctgcgcaa

        19051  ggaccgcccc gaggagctct ccgccgtcac cggcctggcc cgcgcccacg

        19101  tccgcggcgt cacggtccgc tgggccggcc tcttcgacgg caccggcgcg

    20  19151  cgccgcgccg acctgcccac ctaccccttc cagcaccagc ggttctggcc

        19201  gaccgcggcc cgcgccgccc aggacgtcac cgccgcggga ctgggcgccg

        19251  ccgaccaccc gctgctcggc gccaccgtcg aactcgccga cggggccggc

        19301  tacttgttca ccagccggct ctccgtccgg acccacccct ggctcgccga

        19351  ccacggggtc cagggccggg ccctgctgcc cggcaccgcc ttcgtcgaac

    25  19401  tggccgtccg cgccggcgac gaggccggct gcgaccgcgt cgaggaactg

        19451  accctggccg cccccctggt gctgcccgag cgcggcggcg tccaactcca

        19501  ggtccgcgtc ggcgcccccg acgccgccgg ccgccgcacc ctcggcatct

        19551  tctcccgcgt cgaggacggc ttcgacctgc cctggtcgca acacgccacc

        19601  ggcgtcctga ccgccggcgc cggcgccccc gaccccacct tcgacgccac

    30  19651  cgtctggccc cccagcggcg ccgaacccgt cgacctcacc ggcgcgtacg

        19701  agcgcctggc cgcactcggc ttccagtacg ccccgccctt ccagggcctg

        19751  cgcgccgcct ggcgccgcga caccgaggtc tacgccgaag tggccctgcc

        19801  cgacggcgcg gacaccgacc ccgccgcctt cggactgcac ccggccctgc

        19851  tggacgccgc acaacacgcc gccgcctacg ccgacctcgg cgccatcagc

    35  19901  cgcggcggcc tgccgttcgc ctgggaaggc gtctcgctcg ccgccgccgg

        19951  cgccaccacc gtccgcgccc ggatcgcccc ggccggcgag gacaccgtca

        20001  ccatcgccgt ctacgacgcc gccggcggca ccgtgctgtc cgtcgactcc
```

```
20051    ctggtctccc  gcgaggtccc  cgccgacgca  cccggcgccg  ccggcaccgt
20101    ccaccgcgac  tccctcttcc  acgtcgagtg  gaccccgctc  cagggccgcc
20151    cgggccccgc  accggccacc  gtcgccgtcc  tcggccccga  cccggacgcc
20201    ctcgccgaca  ccctccgcgc  caccggcatc  cggaccaccg  cccccgcga
20251    cctggccgcc  ctcgccgacg  ccgaagggcc  cgtccccgac  ctggtcgtca
20301    ccaccctcac  caccaccccg  ggcgcccccg  tccccgacgc  cgcgcacgcc
20351    accaccgccg  ccgtcctcgc  cctcgcccaa  cagtggctcg  ccgacgaccg
20401    cttcgccgac  gcccgcctgg  tcctcgtcac  ccgcggcgcc  accgacggca
20451    ccgaccccgc  cgccgcggcc  gccggcggcc  tgatccgcac  cgcccgcacc
20501    gagaaccccg  gccgtttcgc  cctcctcgac  ctcgcccccg  acaccggccg
20551    gcccgacccc  gagaccctgg  ccaccgccct  ggccgccagc  cacgacgagc
20601    ccgacctcgc  cgtccgcggc  accgacgtgc  acgccgcccg  cctggcccgt
20651    gtcccgctcg  ccaccgaacc  caccacctgg  aacccggacg  gcaccgtcct
20701    gatcaccggc  ggcaccggcg  gcctgggcgc  ggtcctcgcc  cgccacctgg
20751    tcgccaccca  cggcgtccgc  cacctgctgc  tcgccagccg  ccgcggcccg
20801    gccgccgacg  gcgccgacga  cctgacggcc  gaactcaccg  ggctcggcgc
20851    caccgtccac  atcgccgcct  gcgacgtcgc  cgacccggcc  gccctcgccg
20901    acctgctcgg  caccgtcccg  gccgggcacc  cgctcaccgt  cgtcgtccac
20951    accgccggcg  tcgtcgacga  cggcgtcctc  ggctccctca  ccccgcagcg
21001    cctggacacc  gtcctgcggc  ccaaggccga  cgccgcctgg  cacctgcacg
21051    aggcgacccg  ccacctcgac  ctggacgcct  tcgtcctctt  ctcgtccgtc
21101    gccgccaccc  tcggcagccc  cggacaggcc  aactacgccg  ccggcaacgc
21151    cttcctggac  gccctcgccg  cccggcgcgc  cgccaccggc  ctgcccgcca
21201    cctccctcgc  ctggggcccg  tggacccaga  gcgtcggcat  gacaagcagc
21251    ctgtccgacc  tcgacgtcga  gcgcatcgcc  cgctccggca  tgccccgct
21301    gaccctggaa  cagggcaccg  ccctcttcga  cgcggccctg  gccgccgggc
21351    ccgccgccct  cgccccggtc  cgcctcgacc  tgcccgtcct  gcgcacccag
21401    ggcgacatcg  ccccgctgct  gcgcggcctg  atccgcaccc  ccgtgcggcg
21451    caccgccgcc  caggtctcgc  agaccgccga  cggcctcgcc  cagcggctcg
21501    ccggcctcga  cgccgcgccc  cgccgggaag  ccctcctgga  actcgtccgc
21551    acccagatcg  cccaggtcct  cggccacgcg  gacgccaccg  aggtggagac
21601    cggccgccag  ttccaggacc  tcggcttcga  ctccctcacc  gccgtcgaac
21651    tccgcaacgc  cctgaacacc  gccaccggcc  tgcggctgcc  cgccaccatg
21701    gtgttcgact  acccgacacc  acacgccctc  gccgaccacc  tgcgcgacga
21751    actcctgggc  accgaggccg  agtcgaccac  cgccgtcccc  gtgccgaccc
21801    gtaccgccgg  caccgacgac  ccgatcgtca  tcgtcggcat  ggcctgccgc
21851    taccccggcg  gcatcgcctc  acccgaggac  ctctggcgcc  tggtcagcca
```

```
         21901  gggcgccgac gccactggcc cgttccccac caaccgcggc tgggacctgg
         21951  acaacctcta cgaccccgac cccgaccgcc cgggccgcac ccacgtccgc
         22001  gccggcggct tcctgcacga cgccggctcc ttcgacgccg acttcttcgg
         22051  gatgagcccg cgcgaggcga tggccaccga ctcccagcag cgcctgctgc
      5  22101  tcgaactctc ctgggaagcc gtcgaacgcg ccggcatcga ccccgcctca
         22151  ctgcgcgact ccggcaccgg cgtcttcgcc ggcgtcatgt acaacgacta
         22201  cggcaccacc ctgaccggcg acgagtacga ggcgttccgc ggcaacggca
         22251  gcgccccgag cgtcgcctcc ggccgcgtct cctacaccct cggcctggaa
         22301  ggcccggccg tcacggtgga caccgcctgc tcttcctctc tggtcgccct
     10  22351  gcactgggcg cgcaggcgt tgcgggcggg ggagtgctcg ttggcgttgg
         22401  ccggtggtgt gacggtgatg tcgacgccga gcacgttcgt ggagttctcg
         22451  cggcagcggg gtctggcgcc tgatggtcgt tcgaaggcgt tcgccgaggc
         22501  cgcggacggc gtggcctggt ccgagggcgt cggcatgctg gtcctggagc
         22551  ggcagtcgga cgcggtgcgc aacggtcacg agatcctggc cgtggtgcgc
     15  22601  ggctcggcgg tcaaccagga cggtgcgtcc aacggtctga ccgcgcccaa
         22651  cggcccgtcc cagcagcggg tgatccgtca ggcgttggcc agtggcggcc
         22701  tgtccacggc cgacgtggac gccgttgagg cgcacggcac gggtacgacg
         22751  ctcggtgacc cgatcgaggc ccaggcgctc ctggccacct acggtcgcga
         22801  ccgcgacccc gagaacccgc tgctgctcgg ctcgatcaag tccaacatcg
     20  22851  gtcacaccca ggcagcggcc ggtgtcgccg gtgtcatcaa gatggtcatg
         22901  gcgatgcggc acggcgtgct gccgcagacc ctgcatgtcg acgcgccgtc
         22951  ctcgcacgtc gattggagcg tcggcgccgt cgaactgctc accgagcaga
         23001  ccgcctggcc ggagaccggc cgggcccgtc gcgccggtgt ctcctccttc
         23051  ggcatcagcg gcaccaacgc ccacgtcgtc atcgagcagt ccccgaccgc
     25  23101  cgtccccgcc acgcccgcgt ccgccgaccg gtccgtcgag gaaccgccgg
         23151  ccgtcccctg ggccctgtcc ggcaagaccc ccgacgccct ccgcgaccag
         23201  gccgcccgcc tcctcgccca cgtcgaggcc cacccccgcac tgcgcccccgt
         23251  cgacatcagc tactccctga tcgccacccg caccgccttc gaccaccgcg
         23301  ccgtcgtcct cggcaccgac cgcgccgagg ccctgcgcgc cctcaccgcc
     30  23351  ctcgccgccg gcgagaccga cccggccgcc ctcaccggca ccgtccgcac
         23401  cggccgcacc gccttcctct tctccggcca gggctcccaa cggctcggca
         23451  tgggccgcgt cctctacgag cggttccccg ccttcgccga agccctcgac
         23501  accgtcctca ccgccctcga cgcggaactc ggccacccccc tccgcgacat
         23551  catctggggc gaggacgctc aactcgtcga ccggaccggc tacacccaac
     35  23601  ccgccctgtt cgccatcgag gtggcactct tccgcctcct ggaagcctgg
         23651  ggcatcacac cggacttcgt ggccggccac tccatcggcg agatcgccgc
         23701  cgcacacgtc gccggcgtgc tctccctcgg cgacgcctgc cgcctcgtcg
```

```
     23751   tggcccgcgc   cgtgctgatg   cagtcgctgc   ccgaaggcgg   cgcgatgatc
     23801   gccgtccagg   ccaccgagga   cgaggtcctg   cccctcctca   ccgacgacgt
     23851   ctcgatcgcc   gccgtcaaca   gcccgacctc   cgtcgtcgtc   tccggctacg
     23901   agaacgccac   cctcgccgtc   gcccggcact   cgccgacca   gggccgccgc
  5  23951   accacgcggc   tgcgcgtcag   ccacgccttc   cactcgccgc   tgatggcgcc
     24001   gatgctcgac   gacttccgcg   ccgtcgtcga   gagcctcacc   ttcaccgccc
     24051   ccacgacccc   cgtcgtctcc   aacctgaccg   gcgaactggc   cccggccgag
     24101   gcgctctgct   cggccgacta   ctgggtccgg   cacgtccgcg   aggcggtccg
     24151   cttcgccgac   ggcatccgca   ccctcgccga   ccgcggcgtc   accaccttcg
  10 24201   tcgaactcgg   ccccgacagc   gtgctgtccg   ccatggccca   ggagtccgcc
     24251   cccgaaggcg   ccggcaccat   cccgctcctg   cgccgcgacc   ggcccgagga
     24301   acaggccgtc   ctggccgccc   tctgccacct   ccaggtgctc   ggcgtcgagg
     24351   ccgactggtc   cgccaccttc   cgcggcctcg   accccgtccg   cgtcgacctg
     24401   ccgacctacg   ccttccagca   ccgctggttc   tggcccgccg   cccgacccgc
  15 24451   ccgccccgac   gacgtccgcg   ccgccggcct   gggcgccgcc   gaacacccc
     24501   tcctcggcgc   cgccgtgcaa   ctccccgacg   acgacggcgc   actcttcacc
     24551   ggccgcctct   ccctgcgcac   ccaccgtgg   ctggccgacc   acaccgtcct
     24601   gggcaccgtc   ctgctcccgg   gcaccgcact   ggtggaactc   gccgtccgcg
     24651   cgggcgacga   gaccggcagc   ggccacctcg   aagaactcac   cctcgccgcg
  20 24701   cccctgaccc   tccccgagga   cggcgccacc   ctcctccagg   tccgcgtcgg
     24751   atccgccgac   gacaccggcc   gccgcaccgt   caccgtccac   gcccgccccg
     24801   acgacaccgc   cgaccgcacc   tggacgctgc   acgccaccgg   tgtgctcgcc
     24851   accacgccac   cggccgccgc   ggcgttcgac   accacggtct   ggccgcccgc
     24901   cgacgccgaa   cccctcacca   ccgacgactg   ctacgcacac   ttcaccaccc
  25 24951   accgcttcgc   ctacggcccc   gccttccagg   gcctgcgggc   cgcctggcgc
     25001   gccggcgacg   tgctgtacgc   cgaggtcgcc   ctgccggagt   ccgccaccga
     25051   cgaagcggcc   gccttcggcc   tgcacccggc   gctcctggac   gccggcctgc
     25101   acgccgcgct   cctcgccgac   gaccgcgaca   ccggactccc   gttctcctgg
     25151   gaaggcgtca   ctctgcacgc   ctccggcgcc   accgcgctac   gcgtccggct
  30 25201   cgccccgaac   ggccccaacg   gcctgtccgt   caccgccgcc   gacccggccg
     25251   gcaaccccgt   cgccaccgtc   accgcctgc   tcgcccgccc   cctggacgcc
     25301   gagcagttga   ccatccacag   cgccctgacc   cgcgacgcgc   tcttccacct
     25351   ggactggacc   ccggtcccgc   ttcccgacac   cgccaactcc   gcgccgccgg
     25401   ccctcctcgg   cccggacacc   gccgtgctcg   ccgacgccct   cggcgacccg
  35 25451   gccgtcgcac   gccacgcaac   cctcgacgac   ctcctggccg   gggacaccac
     25501   cccgccccgc   acggtcctcg   tccccctcgg   cgccccactc   gacggcgaca
     25551   ccgcgcagca   cgcgcacgcc   ctcacccgca   gcgcgctgac   cctcgtccag
```

```
      25601  cagtggctcg ccaccgaccg cctcgccgac tcccgcctgg tcttcgtcac
      25651  ccacggagcc gtcgccaccg acgacgcgcc ccccaccgac ctggccgccg
      25701  ccgcggtctg gggcctgatc cgctccgcgc agaccgagaa ccccggcacc
      25751  ttcaccctcc tcgacctcga caccgagccc gactcgacca ccgcgctcag
   5  25801  ccgcgccctg accctcgacg aaccacagct cctcctccgc gccggccgcg
      25851  cccgcgccgc ccgcctcacc cgcacccccg cccccaccac caccacccac
      25901  acgccgtggt ccgcggacgg aacggtgttg gtgacgggtg gtacgggtgg
      25951  tctgggtggg ttggtggccc ggcatctggt gcggtcgtgt ggggtgcggc
      26001  atttgttgtt gaccagtcgt tctggtgtgg gtgctgcggg tgcggccggg
  10  26051  ttggtcgcgg agttggagtc gttgggcgcg cgggttgtgg ttgcggcgtg
      26101  tgatgtgggt gatggctcgg ctgttgcgga gttggttgcc ggtgtgtcgg
      26151  agtcgtatcc gttgtctgcg gtggtgcatg cggctggtgt gttggatgac
      26201  ggtgtggtgg gttcgttgac gccggagcgg ttggctgcgg tgttgcgtcc
      26251  gaaggtggat ggtgcgtgga acctgcatga ggcgacgcgt ggtctggatc
  15  26301  tggacgcgtt tgttgtcttc tcgtctgttg cgggtgtgtt cggggggtgcg
      26351  ggtcaggcca actatgcggc gggtaatgcg tttttggacg cgttgatggt
      26401  tcatcgggtg gctggtgggt tgcctggtgt gtcgttggcg tggggtgctt
      26451  gggatcaggg tgtggggatg acggcggggc tgacggagcg ggatgtccgt
      26501  cgtgctgctg agtcgggtat gccgttgttg acggttgatc agggtgtggc
  20  26551  gttgttcgat gcggcgttgg cgacggggag tgccgcgttg gtgccggtcc
      26601  gtctggacct ggccgcactg cgcacccggg cgacatcgc accgctcctc
      26651  cgcggcctcg tccgcgcacc gctgcgccgc accgcggcca ccggcctcgc
      26701  caccggcgcg gacaccggcc tcgtccaacg gctcggccga ctcgaccacg
      26751  cccaacgcca cgaggcactg ctcgacatgg tccgcagcag cgccgcgctc
  25  26801  gtcctcggcc acgccgacgg caacgccatc gacgccgaac gcgccttccg
      26851  cgacctcggc ttcgactcgc tcaccgcggt cgaactccgc aaccgtctgc
      26901  gcaccgccac cggcctgcac ctgtcggcca ccatggtctt cgaccacccc
      26951  accctgtccg ccctcgcgga gcacctgcgg gacgagttgt tcggcgcggt
      27001  cgagagcgag gtgcgggtgc cggtccaggc actgccgccg accgccgacg
  30  27051  atcccatcgt ggtggtgggc atggcctgcc gtttccccgg tggtgtgacc
      27101  tcgcccgagg acctgtggcg cctggtcgac gacggcaccg acgccatcac
      27151  caccttcccg accaaccgcg gctgggacct ggacaacctc tacgacccgg
      27201  accccgagca cttcggcacg tcgtacaccc gctccggtgg cttcctgcac
      27251  gaggcggggg agttcgaccc ggcgttcttc ggaatgagcc cgcgtgaggc
  35  27301  gctggcaacc gactcccaac agcgtctcct gctggaatcc tcctgggagg
      27351  cgatcgagcg ggccggcatc gacccgctga ccctgcgcgg cagcgccacc
      27401  ggcgtcttcg ccggcgtgat gtacagcgac tacgggagca tcctcggcgg
```

SUBSTITUTE SHEET (RULE 26)

```
27451    caaggagttc gagggcttcc aaggccaggg aagtgcgggc agcgtggcct
27501    cgggccgcgt ctcctacgcc ctcggcttcg agggcccggc cgtcacggtg
27551    gacacggctt gctcttcctc tctggtcgcc ctgcactggg cggcgcaggc
27601    gttgcgggcg ggggagtgct cgttggcgtt ggccggtggt gtgacggtga
27651    tgtcgacgcc gagcacgttc gtggagttct cgcggcagcg gggtctggcg
27701    cctgatggtc gttccaaggc gttcgccgag gccgcggacg gcgtcggctg
27751    gtccgagggc gtcggcatcc tcgtcctgga gcgccagtcg gacgcggtgc
27801    gcaacggcca cgagatcctc gccgtgatcc gcggctcggc ggtcaaccag
27851    gacggtgcgt ccaacggcct gaccgcgccc aacggcccgt cccagcagcg
27901    cgtcatccgt caggcgttgg ccagtggcgg cctgtccacg gccgacgtgg
27951    acgccgtcga ggcgcacggc acgggtacga cgctcggtga cccgatcgag
28001    gcccaggcgc tcctggccac ctacggccgt gaccgcgacc ccgagaaccc
28051    cctgtggctg ggctccctga agtccaacat cgggcacacc caggcagcgg
28101    ccggtgtcgc cggtgtcatc aagatggtca tggcgatgcg gcacggcgtg
28151    ctgccgcaga ccctgcatgt cgacgcgccg tcctcgcacg tcgattggag
28201    cgtcggcgcc gtcgaactgc tcaccgagca gaccgcctgg ccggagaccg
28251    gccgggtccg tcgcgccggt gtctcctcct cggcatcag cggcaccaac
28301    gcccacgtca tcgtggaaca gccggcgctc gtcgaaagcc cggccgcgga
28351    gccgagcgga cgcgaacccg gcgtcgttcc gctgccgctg tccggaaagt
28401    cccccgaggc cctgcgcgac caggccgcac gcctgctggc cgggttggcg
28451    gagcggcccg cgctgcgccc gctcgacctc ggctactcgc tggcgacgac
28501    ccgttcggcg ttcgaccacc gggcggtggt gctcgccacc gaccgcgccg
28551    atgcggtccg cgcgctgacg gcgctcgccg ccgccgacgc ggatctctcc
28601    gccgtcgtcg gcgacacccg cacgggtcgt cacgcggtgc tgttctcggg
28651    tcagggctcg caacgcctgg gcatggggcg tgagttgtac gagcgtttcc
28701    cggtcttcgc cgaggctctc gatgtcgcga tcgaccacct ggacgccgcc
28751    ttgcccgccc aggccagtct gcgtgaggtg atgtggggcg acgatgtcga
28801    gctgctggac gagacgggtt ggacgcagcc ggctctgttc gccgtcgagg
28851    tggccctgtt ccggctggtg gagagttggg gtgtccgtcc ggacttcgtg
28901    gccggtcatt ccatcggtga gatcgcggcg cgcatgtcg tcggggtgtt
28951    ctcgctggag gacgcctgcc gtctggtggc cgcccgtgcg acgctgatgc
29001    aggcgctgcc gaccggcggc gcgatgatcg cgatccaggc cgccgaggac
29051    gaagtcaccc agcacctgac tgacgacgtc tcgatcgccg ccgtcaacgg
29101    cccgacctcc gtggtcgtct ccggggccga gagcgctgcc cgcacggtgg
29151    ccgaccggct cgcggagaac ggccgcaaga cgacccggct gcgggtctcg
29201    catgcgttcc actcgccgtt gatggatccg atgctggcgg agttccgtgc
29251    ggtggccgag ggcctgtcct acgccacccc gaccctcccc gtcgtctcga
```

17

```
      29301  acctgacggg  ccggctggcc  acggccgatg  acctctgctc  ggccgagtac
      29351  tgggcgcgcc  acgtccgcga  ggcggtccgc  ttcgccgacg  gcgtcagcac
      29401  cctggagaac  gagggcgtca  ccacgttcct  ggaactggga  ccggacggcg
      29451  tgctgtccgc  catggcccag  cagtcgctca  ccggcgacgc  cgccaccgtc
   5  29501  ccggccctcc  gcaaggaccg  cgacgaggag  acgtccgcgc  tcaccgccct
      29551  cgcccacctc  cacacggcag  gtctccgcgt  cgactgggcg  gcgttcttcg
      29601  ccggcagcgg  cgccacccgc  gtggacctgc  cgacctacgc  cttccagcac
      29651  gccacctact  ggcccaccgg  caccctgccc  accgcgcacg  ccgcggccgt
      29701  cggcctcacc  gccgccgagc  acccgctgct  gaacggttcg  gtcgaactcg
  10  29751  ccgaaggcga  aggggtgttg  ttcaccggac  ggctgtcact  gcagtcacat
      29801  ccgtggctgg  ccgaccacgc  cgtcatggga  caggtcctgc  tgcccggcac
      29851  cgcactgctg  gaactggcgt  tccgggccgg  cgacgaggcc  ggttgcgacc
      29901  gcgtcgagga  actgacgctc  gccgcaccgc  tcgtcctgcc  cgagcgcggt
      29951  gcggtacaga  cccaggtccg  ggtcggcgtc  gccgacgaca  ccggccgccg
  15  30001  taccgtcacc  gtccactccc  ggcccgagca  cgcgaccgac  gtgtcgtgga
      30051  cccagcacgc  gaccggcacc  ctgaccatgg  gctccgcccc  ggccgacacc
      30101  ggtttcgacg  ccactgcctg  gccgcccgcc  gacgccgaac  ccctcgccac
      30151  cgacgactgc  tacgcgcgct  tcacgacgct  cggcttcgcc  tacgggccgg
      30201  tcttccaggg  cctgcgggcc  gcctggcgcg  ccggtgacgt  gctgtacgcc
  20  30251  gaggtggccc  tggcggagtc  caccggcgac  gaggcgaccg  ccttcggtct
      30301  gcaccccgca  ctgctcgacg  ccgccctgca  cgcctccctc  gtcgcccacg
      30351  agggcgagga  gagcaacggc  ggactgccgt  tctcctggga  gggcgcgacc
      30401  ctctacgcga  ccggcgccac  cgcgctgcgc  gtccggctga  ccccgacggg
      30451  caccgacggc  cgttcggtgg  ccatcgccgt  ggccgacacc  gccggtcgtc
  25  30501  cggtcgccgc  catcgacaac  ctcgtctcgc  gccgggtctc  cggcgaccag
      30551  ttgaccggcg  ccgcgggact  ggcccgcgac  gccctgttca  ccctggactg
      30601  gaaccccgta  ccggagaacc  tcgtaccgga  gaaccctgta  ccggagaaca
      30651  ccggcggggg  ccacgcccag  gaccaggacg  gccggcccgc  cgcggccacc
      30701  gtcgcgctgg  tcggcgcgga  cggcaccgcg  atcgccgccg  acctgaccgc
  30  30751  cgccggcatc  cacaccaccc  tccacccccga  cctcaccacc  ctcgccacga
      30801  ccgacgccga  cgttccgaag  acggtcctca  tccccctcac  cggaaccgga
      30851  accggaaccg  gcaccgggac  tgagtcgacg  gacggaatcg  ggacgggggc
      30901  cgccgagtcg  gacgcgtccg  cccccctcccc  ggccgaggtc  gcccacaccc
      30951  tgtccaccgc  cgcactcgcc  ctcgtccagg  agtggaccgc  acaggagcgc
  35  31001  ttcgccggct  cccgcctggc  gttcgtcacg  accggggcga  cggccgccgg
      31051  cggtaccgac  gtcatggacg  tggccgccgc  cgcggtctgg  ggcctggtcc
      31101  gatccgccca  gtccgaagcc  ccggacacct  tcgtcctgat  cgaccgtgac
```

```
           31151  cccggcccgg  ccggcacgca  cgaccgcaca  gccgccgccg  aacggggcca
           31201  actgctccta  cgggcactgc  acaccgacga  accgcagctc  gccctgcgtg
           31251  acggcggcgt  gctcgccgcc  cgcctggccc  gcttcgacac  cgcggccgcg
           31301  ctcaccccgc  cggccgaccg  ggcctggcgg  ctcgacagca  cggccaaggg
    5      31351  cagcctcaac  ggcctcgccc  tgaccccgta  tccggcggca  ctggcgccgc
           31401  tcaccggcca  cgaggtgcgg  gtcgaggtgc  gtgccgcggg  cctgaacttc
           31451  cgtgacgtgc  tcaacgcgtt  ggggatgtat  ccgggtgatg  atgtcggatc
           31501  gttcggttcg  gaggcggccg  gtgtggtcgt  cgaggtcgga  ccggaggtga
           31551  ccggcctggc  ccccggcgac  caggtcatgg  gcatgatcac  cggcagcttc
    10     31601  ggctcgctcg  ccgtggacga  cgcgcggcgc  ctcgcccgcc  tgcccgagga
           31651  ctggtcctgg  gagacgggtg  cgtcggtgcc  gttggtgttc  ctcaccgcgt
           31701  actacgccct  gaaggagttg  ggtggtctgc  gggcggggga  gaaggtgctg
           31751  gtgcatgccg  gtgccggtgg  tgtcggtatg  gcggcgatcc  agatcgcccg
           31801  gcatgtcggt  gccgaggtgt  tcgccacggc  cagtgagggc  aagtgggacg
    15     31851  tgttgcgctc  gctcggcgtg  gccgacgacc  acatcgcctc  ctcccgcacc
           31901  ctcgacttcg  aggcggcctt  cgccgaagtc gccggcgacc  gcggcctgga
           31951  cgtcgtcctc  aactccctcg  ccggtgactt  cgtcgacgcc  tcgatgcggt
           32001  tgctcggcga  cggcggccgg  ttcctggaga  tgggcaagac  cgacatccgc
           32051  gccgcggact  ccgttcccga  cggcctctcc  taccagtcct  tcgacctcgc
    20     32101  ctgggtggtg  ccggaaacca  tcggcaccat  gctggccgag  ctgatggacc
           32151  tcttccgcac  cggcgcactg  cggccactcc  cggtccgcac  ctgggacgtc
           32201  cggcacgcca  aggacgcgtt  ccgcttcatg  agcatggcca  agcacatcgg
           32251  caagatcgtg  ctcaccctgc  cccgctcctg  gaagcccgag  ggaacggtgt
           32301  tggtgacggg  tggtacgggt  ggtctgggtg  ggttggtggc  ccggcatctg
    25     32351  gtgcggtcgt  gtggggtgcg  gcatttgttg  ttgaccagtc  gttctggtgt
           32401  gggtgctgcg  ggtgcggccg  ggttggtcgc  ggagttggag  tcgttgggcg
           32451  cgcgggttgt  ggttgcggcg  tgtgatgtgg  gtgatggctc  ggctgttgcg
           32501  gagttggttg  ccggtgtgtc  ggagtcgtat  ccgttgtctg  cggtggtgca
           32551  tgcggctggt  gtgttggatg  acgtgtggt  gggttcgttg  acgccggagc
    30     32601  ggttggctgc  ggtgttgcgt  ccgaaggtgg  atggtgcgtg  gaacctgcat
           32651  gaggcgacgc  gtggtctgga  tctggacgcg  tttgttgtct  tctcgtctgt
           32701  tgcgggtgtg  ttcggggggtg  cgggtcaggc  caactatgcg  gcgggtaatg
           32751  cgttttggga  cgcgttgatg  gttcatcggg  tggctggtgg  gttgcctggt
           32801  gtgtcgttgg  cgtggggtgc  ttgggatcag  ggtgtggggga  tgacggcggg
    35     32851  gctgacggag  cgggatgtcc  gtcgtgctgc  tgagtcgggt  atgccgttgt
           32901  tgacggttga  tcagggtgtg  gcgttgttcg  atgcggcgtt  ggcgacgggg
           32951  agtgccgcgt  tggtgccggt  ccgtctggac  ctggccgcac  tgcgcacccg
```

```
      33001  gggcgacatc gcaccgctcc tccgcggcct ggtcaaggcg cccatccgcc
      33051  gcgcagccgc caccacaccc ggcgacaccg gactcgccga gcagctcacc
      33101  cggctccagc gcgccgagcg acgggacacc ctcctcgcgc tcgtccgcga
      33151  ccaggccgcg atggtcctcg gccacacctc gggcgacggc gtcgacccgt
   5  33201  cccgcgcctt ccgcgacctc ggcttcgact cgctcaccgc ggtcgaactc
      33251  cgcaaccgca tcggcgcggc caccggcctg cggctaccgg ccacggccgt
      33301  cttcgactac cccaccgccg atgccctcgc cgcacacctg ctcaccgaac
      33351  tgctcggccc cgacgccgag tcggaccccg acgagcccgg cgaccccacc
      33401  gcgggaccga ccgacgaccc catcgtcatc atcggcatga gctgccgctt
  10  33451  ccccggcgac atcggctcgc cggaggacct gtggcgcctg ctcggcgacg
      33501  gcgccgacgt cgtcaccgac ttcccgacca accgcggctg ggacctggac
      33551  aacctctacg accccgaccc cgcgcacgcc ggcacctcgt acgcccgcac
      33601  cggcggtttc ctgcacgacg ccgccgactt cgacgccgac ttcttcggca
      33651  tgagcccccg cgaggccatg gccacggact cccagcagcg cctgctgttg
  15  33701  gagtcctcgt gggaggcgat cgagcgggcc ggcatcgacc cgctgaccct
      33751  gcgcgacagc cgcaccggcg tcttcgccgg cgtcatgtac agcggctacg
      33801  gcacccgcct cgacggcgcc gaattcgaag gcttccaggg gcagggcagc
      33851  gcactgagcg tggcctccgg ccgggtctcc tacaccttcg gcttcgaagg
      33901  cccggccatg acggtcgaca ccgcctgctc ctcctcgctg gtcgccctgc
  20  33951  acctcgccgc acaggcactc cgcggcggtg agtgcaccct cgccctcgcc
      34001  ggtggtgtca ccgtgatgtc catcccggac accttcatcg agttctcccg
      34051  gcagcgcgga ctggcccccg acggccgctc caagccgttc tccgagtccg
      34101  ccgacggcgt cggctggtcc gagggcgtcg gaatgctgct cctggagcgc
      34151  cagtcggacg ccgtgcgcaa cggccaccag atcctggccg tggtgcgcgg
  25  34201  ctcggcggtc aaccaggacg gtgcgtccaa cggcctgacc gcgcccaacg
      34251  gcccgtccca gcagcgggtg atccgtcagg cgttggccag cggcggcctg
      34301  tccacggccg acgtggacgc cgtcgaggcg cacggcacgg gcaccacgct
      34351  cggtgacccg atcgaggccc aggccctcct ggccacctac ggccgcgacc
      34401  gcgaccccga aacccgctg ctgctcggtt cgatcaagtc caacctcggc
  30  34451  cacacccagg cagccgccgg tgtcgccggc gtcatcaaga tggtcatggc
      34501  gatgcggcac ggcgtgctgc cccgcagcct gaacatcacc gagccgtcct
      34551  cgcacgtcga ttggagcgcc ggcgccgtcg aactgctcac cgagcagacc
      34601  gcctggccgg agaccggccg ggcccgtcgc gccggtatct cctccttcgg
      34651  catcagcggc accaacgccc acgtcatcct ggagcagccg gaggccgcgc
  35  34701  ggcactcggc gccggaagaa gccgacacgg cggaggcagc cgccaaggcg
      34751  ccggccaccg cgcacctgcc cgtaatgccg tgggcactgt ccggcaagac
      34801  gccggaggcc ctgcgtgccc aggccgcacg cctcctcgcc cacctccagc
```

```
      34851   agcgccccga actcgcaccc gccgacatcg ccctgtccct cgccacccag
      34901   cgctcccagt tcacccaccg ggcagtcgtc ctgagcaccg accgtgacga
      34951   ggcgacccgc gcgctgtccg ccctcgccac caccgccgcg tccgacccct
      35001   cggccctcac cggcacggtc accatgggac gttgcgcggt gctgttctcg
  5   35051   ggtcagggct cgcaacgtct gggcatgggg cgtgagttgt acgagcgttt
      35101   cccggtcttc gccgaggctc tcgatgtcgt gatcgatcac ctggacgccg
      35151   ccttgcccgc ccaggccggt ttgcgtgagg tgatgtgggg cgacgatgtc
      35201   gagttgctga acgagacggg ttggacccag cccgcgctct cgccatcga
      35251   ggtggcgctg tttcggctgg tggagagttg gggtgtccgt ccggacttcg
 10   35301   tggccggtca ttccatcggt gagatcgcgg cggcgcatgt cgtcggggtg
      35351   ttctcgttgg aggacgcgtg ccgtctggtg gccgcgcggg cgacgctgat
      35401   gcaggcgttg ccggccggtg gcgcgatgat cgcggtccag gcgaccgagg
      35451   acgaagtcat cccgcacctg accgacgagg tggcgatcgc ggccgtcaac
      35501   ggcccgacct ccgtggtgat ctcgggcgca gaagaggcca cgcagaccgt
 15   35551   ggcacaacac ttcgccgacc aggggcgccg gacgaccgcg ctgcgggtct
      35601   cgcatgcgtt ccactcgccg ctgatgatgc tggcggagtt ccgtgcggtg
      35651   gccgagggcc tgtcctacgc caccccgacc ctccccgtcg tctcgaatct
      35701   gacgggccag gtggccacgg ccgacgaact ctgctcggcc gagtactggg
      35751   tgcgccacgt ccgtgaggcg gtccgcttcg ccgacggtgt gacggccctc
 20   35801   gaagccgagg gcgtgcggac cttcctggaa ctcggcccgg acggcgtcct
      35851   cgccgccatg gccaggaaa ccgtcgccga cgacacggtc accgtccccg
      35901   tcctccgcag gaacatgccc gaggaacgga ccctgctcac cgcactcggc
      35951   cggctccaca ccaccggaac cccgatcgac tgggccgccc tcctggcccc
      36001   gaccggcgcc cgcccggtgg acctgccgac atacgcgttc caacaccgtc
 25   36051   ccttctggcc ctccggcccc cgcgacaccg cggatgccgc cgccgtcggc
      36101   atcgccggcg cgagccaccc gctcctcaac ggcatcgtcg aactcgccga
      36151   cgaagagggc ctgttgttca ccggacggct gtcactgcag tcgcatccgt
      36201   ggctggccga ccacgccgtc atgggacagg tcctgctgcc cggcaccgca
      36251   ctgctggaac tcgccctgcg cgccggcgac gaggtcggct gtgaccacgt
 30   36301   cgaggaactg acgctcgccg caccgctcgt cctgcccgag cgcggcgcgg
      36351   tacagaccca ggtccgggtc ggcgtcgccg acaccaccgg cgccgcacc
      36401   gtcacgatcc actcgcgtcc cgcacgcgcc acgaccaccg acagtgacac
      36451   ccacaccggc accgacaccc cgtggaccca acacgccacc ggcgtcctcg
      36501   tcgccggcct gccggcgacg gcaaccgtcc cgttcgatgc caccgtgtgg
 35   36551   ccgccggcgc acgccgaacc cgttgacctg gcggacttct acgcgtcccg
      36601   ggccggcgaa ggattcggct acgggcccgc tttccagggc ctgcgagcgg
      36651   cctggcgccg cgacggcgag gtgttcgccg atgtcgcact gccggaggcc
```

```
36701   ggccgtaccg  aagccgaggc  gtacgggctg  catccggcac  tgctcgacgc
36751   cggactgcac  gcagcctggc  tcgtcgcccc  ggacggggag  cccacacgga
36801   cgggcagcgt  gccgttctcg  tggcgcggcg  tcttcctggc  cgcttccggt
36851   gcctcctcgg  tccgcgtccg  actcggccgc  gactccgacg  gaacgctgag
36901   cctggccatc  gccgacacca  ccggtgcacc  ggtcgcgtcc  gtacaggccc
36951   tctccatgcg  caccgtctcg  gtgacggccc  tgagcgccac  cgcgggcctc
37001   gcccgcgacg  cgctgttccg  cctggactgg  gcctcggccc  cggagccggc
37051   gtgccagccg  gacgacacgg  tgaccgtgat  cccggcggtc  gcggtcgtcg
37101   gtacggaaac  ctccgaactc  acctccgagc  tcaccgcggc  cctgcgtgcg
37151   gcgggcgccg  acgtcgacgt  ccgcacgacc  ctgtcgaccg  atgaacccgc
37201   gcccgcgctg  atcgcgctgc  cgctcgtcgc  ctccgaccag  accggcaccg
37251   cagaagcggc  acccgtcccg  gcggccgtgc  acgacctcac  ccgccgagcc
37301   ctggccctcg  tacagacccg  cctgcaagag  cagcacttcg  cggacacgaa
37351   gttcgtcttc  gtgacccgtg  gtgcgacggt  cgggcgtgat  gtggctgctg
37401   ctgcggtgtg  gggtctggtg  cgttcggcgc  agtcggagaa  tccgggttgt
37451   tttgctctgg  tcgatctgga  tccggatggt  gcggtgggtg  cggctgcgct
37501   cgtcgctgcg  ttggtcagtg  gtgagccgca  gcttgcggtg  cgcggtgatg
37551   tgttgcgggt  cgcgcgtctg  gtgcggcggc  cgctcaccga  ggtcggtgcg
37601   ggtgctgatg  gcaccgggga  tggcgtcggg  ggtggctctg  gtgtgtcgtt
37651   ctcgggtgag  ggtgcggtcc  tggtcactgg  tggtacgggt  ggtctgggtg
37701   cggtgttggc  gcgtcatctg  gtggccgagt  atggggtgcg  ggatctgctg
37751   ttggtcagtc  gcagtggtga  acgtgccgtg  ggtgctgggg  agttggtggc
37801   ggagcttgcg  ggtgtgggtg  cgcgggtgcg  ggtggttgcg  tgtgatgtga
37851   ccgatcgtgc  cgcggtggtg  gagttggttg  gcgggcatgc  ggtgtccgcg
37901   gtggttcatg  cggctggtgt  gctggatgac  ggcatggtgg  gtgcgttgac
37951   cggggagcgg  ttgtccgcgg  tgctgcggcc  gaaggtggat  gctgtctggc
38001   atctgcatga  ggcgacccgc  ggcctggatc  tggacgcgtt  cgtcgtcttc
38051   tcctctctcg  ccggggtctt  cggcagtccc  ggccaggcca  actacgcggc
38101   cgcgaacgcc  ttcctggacg  cgctgatgac  gcggcgccgg  gcggagggac
38151   tgcccggcct  gtcactcgca  tggggaccgt  ggtcgctgac  cgatggcaca
38201   tcgggcatgc  tcgcggacgc  cgaggccgat  cgcctgaccc  gttcgggagt
38251   gccaccgctg  accgcggagc  aaggactggc  actgttcgac  gcggccctgg
38301   cgaccggtga  cgccacctgc  gtcccggtcc  gcctggacct  gtcggcactg
38351   cgtgcccagg  gtgaggtgcc  gcccttgctg  cggtccctga  tccgaggccg
38401   ctcgcgccgc  gccgccgccg  cggaatcggc  aaccgccacc  gggctgcggg
38451   aacggctcgt  cggactgaac  ccggtcgagc  gacaggaagt  cctcctggac
38501   ctcgtgcgcg  gccaggtggc  cctggtcctc  ggccacgccg  acgccgacga
```

```
       38551  cgttcatccg gcacgtgcct tcagggagtt gggcttcgac tccctcacct
       38601  cggtggaact gcgcaaccgc ctcaacaccg tcaccggtct gcggctcccc
       38651  gcgaccatgg tgttcgacta tccgaccgtc gaagtgctgg tctcctacgt
       38701  cctggacgag ttgctgggca cggatgccga ggtggcgacc gtgcagccgg
    5  38751  ccgcggttgc ggtggcggac gatccgatcg tcatcgtggg catggcctgc
       38801  cgctaccccg gcggggtcgc ctctccggac gacctgtgga ggctcgtcac
       38851  ggacggcgtc gacgcggtgt cccccttccc gaccaaccgt ggttgggacg
       38901  tcgaatccct ctatcacccg gaccccgacc atctcggtac ctcctacacg
       38951  cgctcgggtg ggttcctgca tgaggcgggg gagttcgatc cggggttctt
   10  39001  cgggatgagt ccgcgggagg cgttggcgac cgattcccag cagcggttgt
       39051  tgttggagtc gtcgtgggag gcgatcgagc gggccggtat tgatccggtg
       39101  agtttgcggg gtagtcggac gggtgtgttc gcgggggtga tgtacagcga
       39151  ttacagcgcg atgttggcga gtccggagtt cgagggtttc cagggcagtg
       39201  ggagttcgcc gagtttggcg tcgggtcggg tggcctacac gttggggttg
   15  39251  gaaggcccgg cggtgacggt ggatacggcg tgttcgtcgt cgttggtggc
       39301  gatgcactgg gcgatgcagg cgttgcgtag tggtgagtgt gggttggcgt
       39351  tggccggtgg tgtgacggtg atgtcgacgc ctgcggtgtt tgtggacttt
       39401  gctcggcagc ggggtttgtc gccggatggt cggtgcaagg cgtttgcgga
       39451  tgcggccgat ggtgtgggct ggtccgaggg cgtcggcgtg ttggtcctgg
   20  39501  agcggcagtc ggacgcggtg cgcaatggtc acgagatttt ggctgtggtg
       39551  cggggttcgg cggtcaacca ggatggtgcg tccaatggtt tgacggcgcc
       39601  caatggtccg tcgcagcagc gggtgatccg gcaggcgttg gccagtggtg
       39651  gtctgacggc cggtgacgtg gatgtggtgg aggcgcatgg tacgggtacg
       39701  acgctcggtg atccgatcga ggcgcaggcg ttgttggcga cgtatgggcg
   25  39751  ggatcgtgag cctgagcggc cgttgttgtt gggttcggtg aagtcgaatc
       39801  tggggcatac gcaggctgct gcgggtgtgg cgggcgttat caagatggtg
       39851  ttggcgatgc ggcatggtgt ggtgccgcgg acgttgcatg tggatgcgcc
       39901  ttcttcgcat gtggactggt ccgagggtgc ggtggagctg ctcagtgagc
       39951  aggcggcttg gccggagacg ggtcgggtgc ggcgggcggg tgtctcctcc
   30  40001  ttcggcatca gcggcaccaa cgcgcacgtg attctggagc gtccggaggc
       40051  cgcgcggcgt ccggtgatgg agacgaacac cgtggagccg tccacggtgc
       40101  cgtgggtgct ttccggcaag acaccggagg ccctgcgcgc ccaggcggcg
       40151  aaactgttgt cgtcgatcga ggaacgcccg gagcttcgcc tcgtcgacgt
       40201  gggaatgtcc ctggtcaccg gccgctcgac ctttgaacac cgtgcggtgg
   35  40251  tactcgccgc cgaccgtgcc gacgccgccc gcgcattgtc ggcgattgct
       40301  gccgacgagg ccgatgctgc cgctgccacc ggccgcgtcg gcgcgggtcg
       40351  tcacgcggtg ctgttctcgg gtcagggtgc tcaacgtctg ggcatggggc
```

```
        40401  gtgagttgta  cgagcgtttc  ccggtcttcg  ccgaggctct  cgatgtcgtg
        40451  gtcgaccacc  tggacgccgc  cttgcccgcc  caggccggtt  tgcgtgaggt
        40501  gatgtggggc  gacgatgcgg  agttgctgaa  cgagacgggt  tggacgcagc
        40551  cggctctgtt  cgccatcgag  gtggcgctgt  tccggctggt  ggagagttgg
    5   40601  ggtgtccgtc  cggacttcgt  ggccggtcat  tccatcggtg  agatcgcggc
        40651  ggcgcatgtc  gccggggtgt  tctcgctgga  ggacgcctgc  cgcctggtgg
        40701  ccgcccgtgc  gacgctgatg  caggcgttgc  cggccggcgg  tgcgatgatc
        40751  gcggtccagg  cgaccgagga  cgaagtcacc  ccgcatctga  ccgacgacgt
        40801  agcgatcgcc  gccatcaacg  ggccgaacgc  actggtcgta  tcgggtgtgg
   10   40851  aggatgccgc  cgtcgagatc  ggggcgcggt  tcgcggccga  ggggcgtcgc
        40901  acgacccgac  tccatgtgtc  gcatgcgttc  cactcgccgt  tgatggatcc
        40951  gatgctggcg  gagttccgtg  tggtggcgga  gggcctgtcc  tatgctgctc
        41001  cgtccctccc  cgtcgtctcg  aatctgacgg  gccaggtggc  cacggccgac
        41051  gaactgtgct  cggccgagta  ctgggtgcgc  cacgtccgcg  aggcggtccg
   15   41101  cttcgccgac  ggggtgacgg  ccctcgaagc  cgagggcgtg  cggaccttcc
        41151  tggaactcgg  cccggacggc  gtcctcgccg  ccatggcagg  agcctcgctc
        41201  accgaatcct  ccctcgcggt  accgctgctc  cgtaaggacc  ggccggagga
        41251  accggcggca  ctcgccgccc  tggcccagtt  gcacatcgcc  ggcgcgcgcg
        41301  tcgattggcc  cgtgctcttc  gctggtgtgg  gtgcggggcg  ggtggagttg
   20   41351  ccgacgtatg  cgttccagcg  tgggtggttc  tggccggttg  gtcgggttgg
        41401  tgttggtggt  gatgtgggtg  ctgtggggct  tgggtctgcg  gggcatccgt
        41451  tgttgggtgc  tgcggtggag  ttggctgcgg  gtgcgggggt  ggtgttgacg
        41501  ggtcgtctgt  cgttgtcgtc  gcatggttgg  ttggctgatc  atgcggtgat
        41551  ggggcgggtg  tttgttcctg  gtacggcgtt  gctggagatg  gtgatgcgtg
   25   41601  ctggtgatga  ggtggggtgt  ggtcgtgttg  aggagctgac  gttggcggcg
        41651  ccgttggtgt  tgcctgagcg  tggtgggtg   cgggttcagg  ttgctgtgga
        41701  tgctcctgat  gctgcgggtc  gtcgtggtgt  ggggtgtat   tcgtgtcctg
        41751  atggtgtggg  tcaggcggtg  tggtcgcagc  atgctgtcgg  tgtgttggcc
        41801  tctggtgtgg  ctgaccaggt  cggtgggttc  ggtgacggtg  gtgtgtggcc
   30   41851  gccgcagggt  gcggtgtcgg  tggatgctga  gggctgctac  gagctgtttg
        41901  cggatgctgg  gttcggttat  ggcccggtgt  tccaggggtt  gcgtgcggtg
        41951  tggcgtcgcg  gcgaggaact  cttcgcagag  gtcgccctgt  cggacgaggt
        42001  tgctgagagc  gctgatacgg  cgaccggttt  cgggttgcac  ccggcgttgc
        42051  tggatgcctc  gctccatgcc  tcgcttctct  cctcccttga  aggtcaatcc
   35   42101  gccgatggtg  ggcctgcgtt  gccgttcgcg  tgggagggtg  tttccctctt
        42151  cgcctcgggt  gcgacggctt  tgcgcgtgcg  gttggcgccg  gcgggtgagc
        42201  atgcggtgtc  ggtgaccgcg  gtggatccga  ccggtgcgcc  ggtgatttcc
```

```
      42251  atcgacgcgc  ttcgtacccg  tcgcctcacc  ctcgatgagg  tcaacgcatc
      42301  tcacacccag  ctgagcgatg  cgctcttcgg  cgtccaatgg  accacggtcc
      42351  cgagcacccc  ggccgccgac  cacccgtcgg  tagccatcat  cggaaccgac
      42401  cacctggggc  tcgccgaagc  gctcagcagt  tcctctgctg  gcgcgaccac
   5  42451  gaccacgacc  gcggccgcgt  acgagagcct  tgacgcgctg  atcgcggcgg
      42501  ggcccgaagt  gtccgtccct  gacgtcacac  tcatcggtct  caccaccgag
      42551  gacgccatcg  ctcagtacgt  gaacgaccac  gacgccacgg  tggccggcca
      42601  aggcaccatc  ggagccggcg  cggcggccgt  agacgcggct  cgccgcctca
      42651  ccgccgaggc  cctgcgcacg  atccaggcat  ggttggccga  cgagcgtctg
  10  42701  gcggcaaggc  gcctggtctt  cgtgacccgc  ggcgcggccg  acgggcagga
      42751  tgtcgccgcg  gcggccgtcc  agggcctggt  gcgctccgcc  cagaccgaga
      42801  acccgggcac  cttcggcctc  ctcgacctcg  acgggaccga  ggcgtcgacc
      42851  gcggtcctcg  gcgaggctct  cacctccgac  gaaccgcaac  tgcttctgcg
      42901  cgacggacac  ctgcacgccg  ccaggctgac  ccgcctggcg  tcgcccgccg
  15  42951  acacggccgt  gcccacggag  tggaacgcgg  acggcacggt  gctgatcacc
      43001  ggtggtaccg  gcgggctcgg  cgcgcagttc  gcacggcacc  tcgtcgacag
      43051  gtacggcgtc  cgcaatctcc  tgctcgtcag  ccggcgcggc  cccgatgccc
      43101  cgggaaccac  ggagttggtc  gccgagctga  cggcgcacgg  tgccgaggtg
      43151  gccgtgcagg  catgtgacgt  ggccgatggc  gatgcggtgg  cggcgttggt
  20  43201  cgccggcgtg  ccggatgagc  acccgctgag  ggcggtcgtc  cacacggccg
      43251  gtgtgttgga  cgacggagtg  atcggctcgc  tcaccgagga  gcggctcgcc
      43301  accgtcctgc  ggcccaaggc  ggatgccgcc  tggcatctgc  acgaagcgac
      43351  ccgcggcctg  gacctggacg  cgttcgtcgt  cttctcctcc  gtcgcagggg
      43401  tcttcggcgg  cgccggccag  gccaactacg  ccgcggccaa  cgccttcctg
  25  43451  gatgcgctga  tggcccagcg  ccgggcagcg  ggcctgcccg  gactctcgtt
      43501  ggcctggggg  ccgtgggacc  agaccggcgg  aatgacgggc  atgctgtcgg
      43551  acgccgaggc  cgaccgcctc  gcccgctccg  gcatcccgcc  gctctccgcg
      43601  gagcagggcc  tcgccctctt  cgacgcggca  ctcgctcttg  ccggaaccag
      43651  cacgccagac  agggcagccg  gcagcgccgc  cgccagcacg  tcgggaaccg
  30  43701  gcgacacgat  cgccatcccg  gccgcggccc  tcgtcgcacc  ggtccggctc
      43751  gacctggcgg  cgctggccgc  gcaggGtgag  gtccccgcga  ttctgcgcgg
      43801  actggtgcgg  acccgcaccc  gccgtacggc  ggccggcggt  tccgtcaccg
      43851  tggccggact  cgtcaaccgc  ctgtccgggc  tcaccgccga  cgagcggcgc
      43901  caggaacttc  tcgaactggt  ccgcactcag  gcagccctcg  tcctcgggca
  35  43951  cgccgatccg  gcgtccgtgg  actccaccgc  acagttccgt  gacctcggct
      44001  tcgactcgct  gaccgccgtc  gagctgcgca  accggctgag  cacggccacc
      44051  ggcctgcgcc  tgaccgcaac  cctggtcttc  gactacccga  acaccgatgc
```

```
44101  cctcgcggag  cacctgcggg  acgagttgtt  cggcgcggtc  gagagcgagg
44151  tgcgggtgcc  ggtccaggca  ctgccgccga  ccgccgacga  tcccatcgtg
44201  gtggtgggca  tggcctgccg  tttccccggt  ggtgtgacct  cgcccgagga
44251  cctgtggcgc  ctggtcgacg  ccggcaccga  cgccatcacc  accttcccga
44301  ccaaccgcgg  ctgggacctc  gaatcgctct  acgacccgga  ccccgcacac
44351  ctcggcacct  cctacacccg  ctccggtggc  ttcctgcacg  aggcggggga
44401  gttcgacccg  gcgttcttcg  gaatgagccc  gcgtgaggcg  ctggcaaccg
44451  actcccaaca  gcgtctcctg  ctggaatcct  cctgggaggc  gatcgagcgg
44501  gccggcatcg  acccgctgac  cctgcgcggc  agcgccaccg  gcgtcttcgc
44551  cggcgtgatg  tacagcgact  acgggagcat  cctcggcggc  aaggagttcg
44601  agggcttcca  aggccaggga  agtgcgggca  gcgtggcctc  gggccgcgtc
44651  tcctacaccc  tcggcttcga  aggtcccgcc  gtcaccgtgg  ataccgcctg
44701  ctcctcctcc  ctggtcgccc  tgcacctggc  agcccaggcc  cttcgggcgg
44751  gtgagtgcac  gctcgcgctc  gccggtggtg  tgacggtgat  gtccacgcca
44801  ggcacgttcg  tggagttctc  gcggcagcgg  ggtctggcgc  ctgatggtcg
44851  ttccaaggcg  ttcgccgagg  ccgcggacgg  cgtcggctgg  tccgagggcg
44901  tcggcatcct  cgtcctggag  cgccagtcgg  acgccgtgcg  caacggccac
44951  gagatcctcg  ccgtgatccg  cggctcggcg  gtcaaccagg  acggtgcgtc
45001  caacgcctg   accgcgccca  acggcccgtc  ccagcagcgc  gtcatccgtc
45051  aggcgttggc  cagtggcggc  ctgtccacgg  ccgacgtgga  cgccgtcgag
45101  gcgcacggca  cgggtacgac  gctcggtgac  ccgatcgagg  cccaggcgct
45151  cctggccacc  tacggccgcg  accgcgaccc  cgagaacccg  ctgctgctcg
45201  gttcgatcaa  gtccaacctc  ggccacaccc  aggcagcggc  cggtgtcgcc
45251  ggtgtcatca  agatggtcat  ggcgatgcgg  cacggcgtgc  tgccgcagac
45301  cctgcatgtc  gacgcgccgt  cctcgcacgt  cgattggagc  gtcggcgccg
45351  tcgaactgct  caccgagcag  accgtctggc  cggagaccgg  ccgggtccgt
45401  cgcgccggtg  tctcctcctt  cggcatcagc  ggcaccaacg  cccacgtcat
45451  cctggaacag  ccggaggccg  tgcagcgcct  ggcaccggga  gcagcagaga
45501  ccgtggagcc  ggtcgccatc  aagccgtcgg  cggaaccgtc  cctggtgccg
45551  tgggcgctgt  ccggcaagtc  acccgaggcc  ctgcgcgccc  aggccgcacg
45601  cctccgtgac  ttcctggcgg  aacggcccga  accgcgctcg  atcgacatcg
45651  gccactcact  ggccgtcaca  cgctcgcagt  tcgaccaccg  cgcgatcgtg
45701  ctggtcgacg  atgcgaaggc  cccggcggac  agcctggccg  ccctcgcggc
45751  cctggcctcc  ggtgtggccg  atcccgccgt  cgtctccgac  gcggtatcga
45801  ccggcggttc  ggcagtgctg  ttcacaggtc  agggtgctca  acgtctgggc
45851  atggggcgtg  agttgtacgg  ccgtttcccg  gtcttcgccg  aggctctcga
45901  tgtcgtggtc  gaccacctgg  acgccgcctt  gcccgcccag  gccggtttgc
```

```
        45951  gtgaggtgat gtggggcgac gatgtcgagt tgctgaacga gacgggttgg
        46001  acccagcccg cgctcttcgc cgtcgaggtg gcgctgttcc ggctggtgga
        46051  gcgttggggt gtccgtccgg acttcgtggc cggtcattcc atcggtgaga
        46101  tcgcggcggc gcatgtcgcc ggggtgttct cgctggagga cgcctgccgt
    5   46151  ctggtggccg cccgtgcgac gctgatgcag gcgctgccga ccggcggcgc
        46201  gatgatcgcg gtccaggcca ccgaggacga agtgaccccg cacctgaccg
        46251  acgaggtggc gatcgcggcc gtcaacggcc cgacctccgt ggtgatctcg
        46301  ggcgcagaag aggccacgca gaccgtggca caacacttcg ccgaccaggg
        46351  gcgccggacg accgcgctgc gggtctcgca tgcgttccac tcgccgctga
   10   46401  tggatccgat gctggcggag ttccgtgcgg tggcggaagg actgtcctac
        46451  gccaccccgt ccctccccgt cgtctcgaat ctgacgggct ggctggccac
        46501  ggccgacgaa ctgtgctcgg ccgagtactg ggtgcgccac gtccgcgagg
        46551  cggtccgctt cgccgacggc atcaccaccc tcgaagccga gggcgtgcgg
        46601  accttcctgg aactcggccc ggacggcatc ctgtccgcgc tggctcagca
   15   46651  gtccctcgcc ggcgaagccg tcaccgtgcc cgtcctgcgc aaggaccgcg
        46701  gtgaggagtc cacggccctg acggcccgag cgcatctcca cacccgcgga
        46751  ctgatcgaag actggcagga cttcttcgct ggtgtgggtg cggggcgggt
        46801  ggagttgccg acgtatgcgt tccagcgtgg gtggttctgg ccggttggtc
        46851  gggttggtgt tggtggtgat gtgggtgctg tggggcttgg gtctgcgggg
   20   46901  catccgttgt tgggtgctgc ggtggagttg gctgcgggtg cggggtggt
        46951  gttgacgggt cgtctgtcgt tgtcgtcgca tggttggttg gctgatcatg
        47001  cggtgatggg gcgggtgttt gctcctggta cggcgttgct ggagatggtg
        47051  atgcgtgctg gtgatgaggt ggggtgtggt cgtgttgagg aactgacgtt
        47101  ggcggcgccg ttggtgttgc ctgagcgtgg tggggtgcgg gttcaggttg
   25   47151  ctgtggatgc tcctgatgct gcgggtcgtc gtggtgtggg ggtgtattcg
        47201  tgtcctgatg gtgtgggtca ggcggtgtgg tcgcagcatg ctgtcggtgt
        47251  gttggcctct ggtgcggctg accaggtcgg tgggttcggt gacggtggtg
        47301  tgtggccgcc gcagggtgcg gtgtcggtgg atgctgaggg ctgctacgag
        47351  ctgtttgcgg atgctgggtt cggttatggc ccggtgttcc aggggttgcg
   30   47401  tgcggtgtgg cgtcgcggcg aggaactctt cgcagaggtc gccctgtcgg
        47451  acgaggttgc tgagagcgct gatacggcga ccggtttcgg gttgcacccg
        47501  gcgttgctgg atgcctcgct ccatgcctcg cttctctcct cccttgaagg
        47551  tcaatccgcc gatggtgggc ctgcgttgcc gttcgcgtgg gagggtgttt
        47601  ccctcttcgc ctcgggtgcg acggctttgc gcgtgcggtt ggcgccggcg
   35   47651  ggtgagcatg cggtgtcggt gaccgcggtg gatccgaccg gtgcgccggt
        47701  gatttccatc gacgcgcttc gtacccgtcg cctcacccty gatgaggtca
        47751  acgcatccca cacccagctg agcgatgcgc tcttcggcgt ccaatggacc
```

27

```
47801   acggtcccga gcacccggc cgccgaccac ccgtcggtag ccatcatcgg
47851   aaccgacccc ttcggcctcg cagacggcct ttcggacgcc ttgcccctgg
47901   tcgaggagcg cggtgacctc gcggcgctcg cagcgtcgga gcacccggta
47951   ccggacctcg tcctcgtccc ggtagcgggc acccggcgca caggcgtacc
48001   tgcggacgcc gaaggacaca ccgacgccgg gacatccgac atgctccgat
48051   ccgtgcgtga ggccaccgca caggtactgg agcagatcca gcagtggttg
48101   gcggacgacc ggttcgaggc ggcgcggctg gtgttcgtga cgcgcggggc
48151   ggtttccgtg ggtgagggcg gcatcgccga cctggcggcc tcggccgtct
48201   ggggtctggt gcggtcggcg cagtcggaga atccgggctg cttcggtctt
48251   ctcgacctcg acctcgacct cgcccttgac tccgaccttg cccccgaggt
48301   cgacatcgag cgcgaccgtg accgcgatcc ggtcggtggg accgtgcagc
48351   ccgcgctcgc cgcggccctg cacgcgaccg ccgacgagcc gcagttggca
48401   ctgcgcggcg ggaccgtgca ggccgcccga ctgacccgaa tccccgcgcc
48451   gcagaccgac cgtgccgaga ccgaccctgc cgagaccgac cgtccggaga
48501   tcgacacccg gcggcccggc acggtgctca tcaccggtgg taccggtggc
48551   ctcggtgggt tgctcgcccg gcacctcgtc gccgagcggg gggtacggag
48601   cctggtgctc gccagccgga gcggtctcgc ggccgaggga gcggagaagc
48651   tggtcgccga cctcgaagcg ctcggtgccg tggtggccgt gcagacgtgt
48701   gatgtggccg atggcgatgc ggtggcggcg ttggtcgccg gcgtgtcgga
48751   cgagtacccg ctgacggcgg tcgtccacac ggccggtgtg ttggacgacg
48801   gagtgatcgg ctcgctcacc gaggagcggc tcgccaccgt cctgcggccc
48851   aaggcggatg ccgcctggca tctgcacgag gcgacccgcg atctggacct
48901   ggacgcgttc gtcgtcttct cctccctcgc cggcgtcctc ggtggcgccg
48951   gtcaggccaa ctacgcggcg gcgaacacgt tcctggacgc cttgatggcg
49001   cagcgtcgcg ccgccgggtt gccgggtgtg tcgctggcgt ggggtccgtg
49051   ggaccgggcc ggcggcatga cggggaccct gtcggacgcc gaggccgacc
49101   gcctcgcccg ctccggtgtt ccgccgatct cggcggagca gggccttgcg
49151   ctgtacgacg cggcgaccgc cggtgagcgg ccgctggtgg tgccggtgcg
49201   gctggacctc gccgcgctcc gcgggctcgg tgatgtcccg gcgctgctgc
49251   gcggactggt ccggacgccc gcgcggcgga ccgcggcggc cggtgcggcg
49301   ccgtcggccg atgtgctcac ccggcagttg gccgggctcg gcgggggcgga
49351   gcaggaggag gtcctgctga ggctggtgcg cggtcaggcc gcggtggtgc
49401   tcgggcacgc cgacggctcg gcgatcggtg cggggcgaca gttccaggag
49451   ttgggcttcg actcgctgac cgcggtggag ttccgcaacc gactcaacgc
49501   ggccaccgga ctgcggctgc cggccaccct gctgttcgac tacccgacgc
49551   cggccgacgt cgtcgggcac ctgcgcggcc ggctcggcac cggggaggtg
49601   tcgggtgcgg gctcggtgct ggcggcgctg gacaaccttg aggcggtgat
```

```
      49651  cgccgggctg  tccctcgacg  acgcggggga  gcaccagttg  gtggccggcc
      49701  ggctggaggt  cctcagggcg  aagtgggcgg  acatgcgaag  cgcggaggga
      49751  gctgtggacg  gcggtgcgga  cgtcgacatc  gaggaggcgt  cggacgacga
      49801  catgttcgcg  ctgctggacg  acgagctggg  gctgaactga  gccgctccgc
5     49851  atgagcagtt  ccgcaccaga  agttccacag  tggcctggtc  cacagtgacc
      49901  tggtccacag  tgagctgttc  cacagcgacc  tgttccgcgg  cggttccgta
      49951  gtgagccgtt  ccgcttaggc  gggaaaccgg  ggccccggtg  gtcggacgct
      50001  tgattccggc  caccgggagc  gtcaaaccgg  cctcttctaa  agaaagggaa
      50051  agaactaggg  aaacacccac  agcccgatca  tgtgcaatga  attccgtggc
10    50101  tcggaaaacc  attcgcgggc  catggagtta  cggtgtgatc  acgtgtccgt
      50151  acggcagtga  cggcgtgcac  gggttgccac  ggcaggtccc  cggccgaggc
      50201  gtctccccgc  tgctgcccgg  cggtgcgtgc  cgcttggttc  gtccctagga
      50251  ggtcccccca  tgaccacgtc  caccgaggag  agcctgtggg  cccggtgctt
      50301  ccatccagca  ccggccgccc  ccgtccggct  cttctgtttc  ccacatgcgg
15    50351  gcggctcggc  ctccttctac  ttcccggtgt  cggcccaact  gtcctcggtt
      50401  gccgaggtgt  tcgccatcca  gtacccgggg  cgccaggacc  ggcgcaagga
      50451  agccggtgtc  agtgacctcg  cgaccttggc  cgaccaggtc  tacgacgcgc
      50501  tgcgccccct  gctgaaggag  cggccgagca  cgttcttcgg  gcacagcatg
      50551  ggcgcgacgc  tggccttcga  ggtggcccgg  cgcttcgagg  ccgacgacgg
20    50601  tgacctggtc  cggctgttcg  cctccgggcg  ccgggccccc  tcccgcgtgc
      50651  gtgaagaggc  cgtgcaccgg  cggtccgacg  acggcatcgt  cgaggagctg
      50701  aagctgctcg  ccggcaccaa  caccgcgctg  ctcggcgacg  aggagatcct
      50751  gcggatgatc  ctgcccgcga  tccgcagcga  ctaccaggcc  atcgagacct
      50801  accgctgccc  gcccgacgtc  accgtccggg  cgccgctgac  cgtcctcacc
25    50851  ggcgaccgcg  acccgaagac  ctccctggac  gaggccgagg  cgtggcgcgg
      50901  ccacaccacc  ggggacttcg  acctcaaggt  gcttcccggt  gggcacttct
      50951  tcgtcagctc  cgaggccccg  gcgatcatcg  atctgctccg  ggcgcacctc
      51001  gccggcaacg  gctagcgggc  gcactgcggc  aggccggcgg  tgccggtctg
      51051  ccgcacggcc  cccgcgccgc  ctgagacggc  accatgccac  ccggcgacgc
30    51101  gcgttgcgtg  cgcctgtgga  gcaccgtccc  gcgtgcgtac  gcgggcgcgc
      51151  ggaggacgtg  ctcggccggc  cggatgtcct  ggaggagacc  attcggaaaa
      51201  cgtcaaatgt  ctgacagctg  ctgctcattt  actccagccc  gcactatatg
      51251  atcttgtggt  ggggcatgta  cggagcgtgc  cactcgttcg  ctcctggtgc
      51301  atccatcccg  accgtccgga  tttgtgtgca  ccggacggcc  gtggtcctgg
35    51351  tacgcgtcgg  cgcgaggtcg  gctcctgttc  tcaccccacg  tggaggaacg
      51401  cgtgatgcgg  aagcagtcgg  gttcttccgg  cttactgacc  acgctggtgg
      51451  gacgcgacga  cgaactgcgc  accctcgccc  ggcacgccgc  ggccgcccgc
```

```
      51501  gacgggcggg  ccggcctggt  cctgctccac  gggcccgccg  gcatgggcaa
      51551  gaccagtctg  ctgcggtcct  tcacggcgag  cgatgtctgc  cgcggcatga
      51601  cggtgctgta  cggcacctgc  ggcgagaccg  tcgccggcgc  cgggtacggc
      51651  ggtgtgcgcg  aactcctcgg  cgggctcggc  ctgagcggcg  gcgacgcccg
   5  51701  gcgctcgccc  ctcctggagg  gcctggcggc  ccgcgcgctg  cccgcactca
      51751  ccgcagaccc  cgccggtccc  gacgccgcca  cgggtgccta  cccggtgctg
      51801  cacggcctgt  actggctcgc  cgcccgcctc  atggcccaac  ggccgctggt
      51851  cctcgtcctc  gacgacgtcc  actggtgcga  cgaacgctcc  ctggcctgga
      51901  tcgacttcct  gctgcgacgc  gccgaggacc  tgccgctgct  ggtcgtcctg
  10  51951  gcctggcgca  gcgaggccga  accggtcgcg  cccgcggtgc  tcgccgacat
      52001  cgccgcccag  cgccgcccca  ccgtgctcgg  cctgcacccg  ctcggccccg
      52051  acgacatagg  cgaaatggtg  cgtcgcgtct  tccggaccac  ggccgcaccg
      52101  tcgttcgtca  gccgggtcgc  cgccgtgtcc  ggcggcaatc  cgctggccct
      52151  cgcccgcctc  ctcgacgaac  tccgcgccga  gggcgtccgg  ccggacgccg
  15  52201  ccggggagcg  ccgggccgcc  gaggtcggca  gtcacgtcct  cgcccgctcg
      52251  gtgcgctgcc  tgttggagcg  ccggccgccc  tgggtgcgcg  gcgtggcccg
      52301  tgccatcgcc  gtactcggcc  cggagtgcac  cgagttgctg  gcggcgctcg
      52351  ccggcgtccc  ggccgcgacc  gtcgacgagg  ccctgttggt  gctgcgcagg
      52401  gccggcatcc  tggccgccga  ccgcgtggac  ttcgtccatg  acgtcgtccg
  20  52451  ctccgcggtg  ctcgacgacg  tcgccccgcc  caccctggcc  gaactgcgca
      52501  ccaacgccgc  gctgttgctg  agcgacgccg  gccgcccctc  cgaggagctc
      52551  gccggccagc  tcatgctgct  gccggtgctc  gaccagccgt  ggatggccgc
      52601  ggtgctgcgc  gacgccgccg  cccaggcgga  gagccgcggc  gccccggagg
      52651  ccggtgtgcg  ctgcctctac  cgggtgttgg  aggtggagcc  ggacaacgtt
  25  52701  gccgtccgca  tccagatggc  ccgcgcgctc  gccgagatca  acccgcccga
      52751  ggcgatgcgc  ctcctcaagg  aagcgctctc  cctcgccggg  gacgtccgca
      52801  cccgcgccca  ggtcgccgtc  cagtacggct  tcacctgcct  cgccgtgcag
      52851  gaatccccgt  ccggggtgcg  gatgctggag  gacgcgctcg  ccgagctgac
      52901  cgccgaactg  ggccccgaac  cagggcccgt  ggaccgggag  ttgcggaccc
  30  52951  tcgtggagtc  cgtgctgctc  atcgtcgggg  ccgacgagaa  ggtgacgatc
      53001  ggtgcggtcc  gcgaccgggc  ggcccggctc  accatgccgc  ccggcgacac
      53051  accggcccag  cggcagatgc  tggccatgac  caccgtgctg  accgcgatgg
      53101  acggccggga  cgccggtcg  gccgtcgacc  aggcccgccg  cgccctgcgc
      53151  gcacccggcg  tcgagctgga  accctggtcg  ctgctgtccg  cctccttcgc
  35  53201  cctctccctg  gccgacgagg  tcgccgacgc  gcagtacgca  ctggacctca
      53251  tgctccagta  cggccaggac  aacgcggcgg  tgtggacgta  cgtcctggcg
      53301  ctctccacgc  gcgccctgct  ccaccacggg  gtgggcgcct  tccccgaggc
```

```
53351  cctggccgac gcccagaccg ccgtcgagat cctcggcgag gagcgctggg
53401  cggacggcgc cgtgctgccc cgtgtcgcgc tggccaccgc cctcgtcgac
53451  cgcggcgagc ccgagcgcgc cgaacacgtc ctcgacggca tcacccgccc
53501  ccgcctcgaa cgcttcgtca tcgaatacca ctggtacctc caggcccgcg
53551  cctacgcccg ctgggtccgc ggggatttcc aaggagccct ggacctcctt
53601  ctcgcctgcg gtcggtccct ggaggagtcg cgcttcagca acccggcgtt
53651  cgtgccgtgg tgggccgatg gcgcggtgct cctggcgacc ctggaccgcc
53701  acgaccaggc gcgcgaactc gccgcatacg gaagcgagtt ggccgagcgt
53751  tgggggacgg cgcgcggcct cggactggcc ttcatggccc agggcgtcgc
53801  cgcacccggc cgcgccggca tcgatcacct caccgaggcg gtctcgctgc
53851  tcgcggactc cccggcccgg gccatggagg cccgggccga acttcttctc
53901  ggacacgccc acctgaagcg cgacgacctg cgggccgccc gggaacacct
53951  gcgcgccgcc gccgacctcg cccagcgctg cggcgccgtg aagctcggcg
54001  tcgacgccag aaaactgctg gtcaccgcgg gtggtcgggt acgcaggatg
54051  accgcctccc cactcgacat gctgaccggg atggaacgca cggtggcgga
54101  cctcgcggtg accggcgcga gcaaccgggc catcgcggaa gccctcttcg
54151  tgactgtaag gaccatcgaa acccatctca cgagcgtcta ccggaagctc
54201  ggggtcggcg ggcgtgcgga gctgtccgcc gtcctggaga ccaggaccgc
54251  cacctccggt cggcagccgc cggcctgggt ctcccaggca cgcggacgcg
54301  cttgagacga acaggacgag aggtagcggt gccccggagc aaggcgcgga
54351  accaaccgac cacctgcacg ccgcagtgcg cacccgacgc ccacggcgac
54401  cccaccatgc tcctcgaatg cggccgggaa cagcggctca tcggcgacct
54451  cctgcaccgc ctcggccagg ggcggccatc ggtgctcagc ctgaccggcc
54501  ggcccgggca cgcccagaac gccctggtcc gctggggcgc gtgccgggcc
54551  aggcacgacg ggctgcgcgt cctgcgagcc caggcgacgc ccgcggaacg
54601  ggaactccgc tacggcgccg ttctccaact gctggccgtc ctcgacggcc
54651  cgcacggcag caccctggac gccgcgatcc gccacgacgg tcccccgcca
54701  ctgcccgtgc ccggcatcga ggaggtgctg cggcgcaccg gcacggcacc
54751  caccctggtc gtggtcgaag acgtccagtg gttggacccg gcctcgctga
54801  cgtggttgca gatcctgctg cgccacctcg ggccggacac cccgctcgcc
54851  gtcctggcca gcagctgcgg tgacaccacg gccttcgaca ccgacccgaa
54901  ggccccgcc gtcccggggc cgccggacac cgtgcccgtc gcgcgcttcg
54951  tggtgcccgc gctcaccgac cgcggggtcg ccgccaccgt ccgcgccgtc
55001  tgcggcaccc ccggcgacga ggagttcatc gccgcgctca cctccgccac
55051  cgccggcaac cccgccatcc tgcgggacgc cctgcgcgcc ttcgtcgacc
55101  acggcctccc cgccgacgcc gaccacctcc cggagctgca cgccctcacc
55151  gctggcgtcg tcggcgacca caccgtgcgc gccctggacg gcctgcccgc
```

```
        53351   cctggccgac gcccagaccg ccgtcgagat cctcggcgag gagcgctggg
        53401   cggacggcgc cgtgctgccc cgtgtcgcgc tggccaccgc cctcgtcgac
        53451   cgcggcgagc ccgagcgcgc cgaacacgtc ctcgacggca tcacccgccc
        53501   ccgcctcgaa cgcttcgtca tcgaatacca ctggtacctc caggcccgcg
    5   53551   cctacgcccg ctgggtccgc ggggatttcc aaggagccct ggacctcctt
        53601   ctcgcctgcg gtcggtccct ggaggagtcg cgcttcagca acccggcgtt
        53651   cgtgccgtgg tgggccgatg gcgcggtgct cctggcgacc ctggaccgcc
        53701   acgaccaggc gcgcgaactc gccgcatacg gaagcgagtt ggccgagcgt
        53751   tgggggacgg cgcgcggcct cggactggcc ttcatggccc agggcgtcgc
   10   53801   cgcacccggc cgcgccggca tcgatcacct caccgaggcg gtctcgctgc
        53851   tcgcggactc cccggcccgg gccatggagg cccgggccga acttcttctc
        53901   ggacacgccc acctgaagcg cgacgacctg cgggccgccc gggaacacct
        53951   gcgcgccgcc gccgacctcg cccagcgctg cggcgccgtg aagctcggcg
        54001   tcgacgccag aaaactgctg gtcaccgcgg gtggtcgggt acgcaggatg
   15   54051   accgcctccc cactcgacat gctgaccggg atggaacgca cggtggcgga
        54101   cctcgcggtg accggcgcga gcaaccgggc catcgcggaa gccctcttcg
        54151   tgactgtaag gaccatcgaa acccatctca cgagcgtcta ccggaagctc
        54201   ggggtcggcg ggcgtgcgga gctgtccgcc gtcctggaga ccaggaccgc
        54251   cacctccggt cggcagccgc cggcctgggt ctcccaggca cgcggacgcg
   20   54301   cttgagacga acaggacgag aggtagcggt gccccggagc aaggcgcgga
        54351   accaaccgac cacctgcacg ccgcagtgcg cacccgacgc ccacggcgac
        54401   cccaccatgc tcctcgaatg cggccgggaa cagcggctca tcggcgacct
        54451   cctgcaccgc ctcggccagg ggcggccatc ggtgctcagc ctgaccggcc
        54501   ggcccgggca cgcccagaac gccctggtcc gctggggcgc gtgccgggcc
   25   54551   aggcacgacg ggctgcgcgt cctgcgagcc caggcgacgc ccgcggaacg
        54601   ggaactccgc tacggcgccg ttctccaact gctggccgtc ctcgacggcc
        54651   cgcacggcag caccctggac gccgcgatcc gccacgacgg tcccccgcca
        54701   ctgcccgtgc ccggcatcga ggaggtgctg cggcgcaccg gcacggcacc
        54751   caccctggtc gtggtcgaag acgtccagtg gttggacccg gcctcgctga
   30   54801   cgtggttgca gatcctgctg cgccacctcg ggccggacac cccgctcgcc
        54851   gtcctggcca gcagctgcgg tgacaccacg gccttcgaca ccgacccgaa
        54901   ggccccgcc gtcccggggc cgccggacac cgtgcccgtc gcgcgcttcg
        54951   tggtgcccgc gctcaccgac cgcggggtcg ccgccaccgt ccgcgccgtc
        55001   tgcggcaccc ccggcgacga ggagttcatc gccgcgctca cctccgccac
   35   55051   cgccggcaac cccgccatcc tgcgggacgc cctgcgcgcc ttcgtcgacc
        55101   acggcctccc cgccgacgcc gaccacctcc cggagctgca cgccctcacc
        55151   gctggcgtcg tcggcgacca caccgtgcgc gccctggacg gcctgcccgc
```

```
       55201  cgaagtcaac  gccgtcctgc  gggccctggc  cgtctgcggc  gacctgctcg
       55251  acttccaccg  agtccgggcc  ctcgccggcg  cgcactcgct  gtccgaggac
       55301  cggatccgca  ccctgctggc  gagcgtcggc  ctgaccgtgt  ccgtcggcga
       55351  caaggtgcac  atccgcttcc  ccgcctccaa  ggcacgcgtc  atcgaggaca
   5   55401  tgcccgccgc  ggagcgcgcc  gatctgtacg  tccgcgcggc  cgaactcacc
       55451  cacagttgcg  gcgtcaacga  cgaggacgtc  gcccatctgc  tgctgcgctc
       55501  gtcgccgctc  ggcgcaccct  gggtcgtgcc  cctgctccgc  cgcggattcg
       55551  ccgccgcgct  cgccgggag  gaccaccacc  gggcctgtgc  ctgcctctcc
       55601  cgcgccctgc  aggaaccccct  cgaccccgg  gaacgcagcc  tgctgacgtt
  10   55651  ggaactggcc  gcggccgaag  ccgtcgcccg  gccggaggcg  ggggatcgac
       55701  gcctggggga  actcgtccgc  agcaccgtcg  cggacaccga  ccccacgtcg
       55751  tccggtgagg  gggtgggggt  ccgcgccatc  gacctggggt  tcgcccgggg
       55801  caacagcgaa  tgggtccgcc  gcaccgcggg  cgaggccctg  ccgtacgccg
       55851  ggccggccga  ccgggaggaa  ctggtcgcgc  tgttctggtt  ggccgccgtg
  15   55901  cgggacgacg  acgcgccgat  gatccccgtg  gtgccccggt  tgcccgaccg
       55951  gccggtgccg  ccggcccagg  ccggcgcccg  tgcctggcag  ctggccacgg
       56001  cgggggagga  cgcggacaag  gccaggaagc  tcgcccggat  cgccctcacc
       56051  ggcggggtga  acgagagcct  gatgatgccg  aaactggcgg  cctgcgccgc
       56101  gctgttcgcc  accgacgaca  cgacgaggc  ggtgcacggc  ctggacacca
  20   56151  tgctcaccgc  cgcccgcagt  gcccacctgc  gcagcatggc  cgcccgcatt
       56201  ttcaacctac  gggcccggat  acacctgtgc  gcggcccggc  tggaggccgc
       56251  cgaacgcgat  ctggacagcg  ccgagcgcgc  cctgccgccg  acgagttggc
       56301  accccgtgc  gctgcccaac  ctgatcgcca  cccgcatcct  cgtcagcatg
       56351  gagacgggcc  gcccggaccg  ggcccgccga  ctcgccgagg  ccccggtccc
  25   56401  cgccggcggc  gaggagggtg  tgtggtggcc  cgccctgctg  ctcgcccgcg
       56451  cccgtgtggc  cgccgacgac  ggtgactggg  aggaggccct  gcggctgtcg
       56501  cgggagtgcg  ggcgctggct  ttttcgccgg  cactgggcca  acccggccat
       56551  gctcagttgg  cgcccgctgg  ccgccgaggc  gtgtctgaag  ctcggcgacg
       56601  tgacggaggc  gcgccggctg  cgggacgagg  agctgttctt  cgccgaccgc
  30   56651  tggggcaccg  cgagcgcccg  cgggatcgcc  cgcctgacga  cgcggcgact
       56701  cttcgacgac  gacggcgacc  gggccgtccg  gcggatccgc  gaggccgccg
       56751  ccctgctccg  cgactcgccc  gcccgcctgg  cctacctgtg  gagccggctg
       56801  agccaggccg  gtgccgagac  ggcccacggc  gacaccgccg  cggccgcacg
       56851  ctcctggcag  gcggtcgccc  ggatgaccgc  cgcccacccc  gccagccgcc
  35   56901  tcgccaccgc  cgcccgcacc  ctgaccgtcc  cgtccgttcc  ggtcgccacc
       56951  gcgccgccca  ccgccgtcgt  cccacccgga  tggcgcgacc  tgtccgaggc
       57001  ggagaaggac  accgtgctgc  tcgccgcccg  cggccacggc  aaccgccaga
```

```
        57051   tcgccgaaca actcgccgtc agcaggcgca ccgtggagct ccggctgagc
        57101   aacgcctacc gcaagctgag gatcggcgga cgcaaggagc tgtacctgct
        57151   cctggaggcg ctggaaggac cggtcgcgga tgcttcttga gcgggagaac
        57201   gaactggccc ggatccgggc cgccctggac gccgcggaag cgggcgactc
    5   57251   ctcgctcctc ctgatcaacg gtcccctcgg cagcgggcgt tcggcgctgc
        57301   tgcgccggat accgagctg gccggcgacg gcacccgcgt cctgcgggcc
        57351   agcgccgcct ggcgggaacg cgacttcccc ttcgggatcg cccgccaact
        57401   cttcgaccac ctgctgtccg gggcgggcgg cgcagggccg gccgaacgca
        57451   ccgccggggc agagcacttc agccgactga tggacaccgg cgaccgccct
    10  57501   accgggaccg gccccgccct ggaggtctcc caggcagtgc tccagggcgc
        57551   ccaggcgctg ctcgccgacg cgtccgcgga gcggcgcctg ctgatcctgg
        57601   tcgacgacct ccagtgggcc gacggcccgt ccctgcgctg gctggcccac
        57651   ctcacccggc ggctgcacgg cctgcgggcg ctgctggtgt gcacgctggc
        57701   cgacggcgac caccggggca ggtaccccct ggtccgggag gtcgccggcg
    15  57751   ccgcgcacac cgtcctgcgc ctggcgccgc tgtcccggga cgccacccgc
        57801   gtcctgctcg ccgggcccca gggccggccg ccgcaggacg cactggtgcg
        57851   cgccgtgtac gaggcgtcca ggggcaaccc gctgttcctg accgccttcc
        57901   ggagcgctct gcgcgccacc ggaaggccgc ccggcggcga ccacttcggc
        57951   gccgtccggg agctgagccc gacggtgctg cgcgatcggc tcgcgggcca
    20  58001   tctgcggatc cagccgcagc cggtgcgcga ggtcgcggtg gcggtggccg
        58051   cgctgggcga ccacagcgat ccggtgctgc tcgcccagct cgccgggggtc
        58101   gatgaaatcg gtttcgccgg tgcccgccgc gcgctggtgg acgccggcct
        58151   gttggcccgg ggacgggacg tccgcttcgt ccacggcgtc gtccgcgatg
        58201   cggtggactc cctgctcacc ctcgacgagc gggaacgctc gcacgacgac
    25  58251   gccgccgatc tgctgtaccg ctgcgggcgg ccggccgagc aggtcgccgg
        58301   ccatctgctg gccgtggtcc acccgggccg gccctggtcg gaggcggtcc
        58351   tgcgctccgc ggcccacaac gcgctgcgcg ccggccggcc cgccgacgcg
        58401   gcccggtacc tgcgccgcgc cctgctgcac caccgcaccc aggacggctg
        58451   ccgcgcccgc atcctggtcg atctggccac cgccgagcgc gccctcgacc
    30  58501   ccgatgcctg tgtacgccac gtcagccagg cggtcgcgct gctggacacc
        58551   tcgcgggatc gggccgccgc cgtgttgcgc atcccgccgt ccctgctcgc
        58601   cgcccccagc ccgtccgccg tcgagttggt gcggcaggcc gccgccgggc
        58651   tcgacgaacc ggggcagcgg gacgaggagg gagccgacga actcgccctg
        58701   cgcctggagg cgtggctgcg gcactccggc cacgagaacc ccgtcgagct
    35  58751   ggcgtcctcg gtggcgcggc tgcggcgcat gggggcacgg ccgccggtgg
        58801   acagcgtcgc cgaacgcgag ttggtcgccg tgctgttgag tgcgggcgcg
        58851   ctcagcggcc ggctcagcgc cgcggagatc gccgacaccg gcaaccgcat
```

SUBSTITUTE SHEET (RULE 26)

```
        58901  cctcgaacgt gagccggcca ccgccgccca tgcccacacc ccgctgccgc
        58951  tggtgatgct ctcgctgttc gtcgccgagt ccgtgcaggg cgtggcctcc
        59001  tggctggcca gcgaacagca cacccggcgc cggtacgcga ccggcgcgga
        59051  cgatgtgctg ctgaccgccg agcgggcctt cgtcctggtg acgcagggcc
    5   59101  gcccggccgc cgcgcgggag cacgtcgaac gcgccctggt catggacgcc
        59151  ggcgactggt cggaacccgc cgtcatgatg ttcgccgcgg tcgccttcga
        59201  actgcgcgac ccggccttga gcgaacgcat cctggaacgg atccgcgacc
        59251  gccggccggc cggactggcg ctcaccgcca ccggtcagat gctccaggcc
        59301  gccgtcgacg tgcacttcgg ccggggggcgg gacgccctgg acacgctgct
    10  59351  ggcctgtggc cgacgcctgg agaccgtggg atggcgcaat tccgcactgc
        59401  tgccctggcg tccgtatgca atcgggctgc accagcggct cggcgagacc
        59451  gatgccgcgc tgcaactcgc cgaggacgag ctgaggtggg cccgggagtg
        59501  gggtgcgacg acgaacctgg gccgggccct gcgcctgaag ggctggctgt
        59551  tgcaggacga gggactggat ctgctgcgtg agagcgtcga gatcctccgc
    15  59601  gcttcgtcct acgcgacgga actggcgcgc accctcgtcg tcctcgggcg
        59651  tcggctgccg ggtgggccgg aggccgaggc ggtgttgcgg gaggccgcgg
        59701  ggatcgccgc cgcctgtggt gttccctggc tcgccgaacg ggccgaactc
        59751  ggtctgggca gtgccatcgt gccgccggtc gccaccctga cccccagcga
        59801  gcggcgagtg gcgtcgctgg tgagccgggg cctgaccaac caggccatcg
    20  59851  cgaccgaact cggtgtgagc tcccgggcgg tggagaagca cctcaccagc
        59901  gcctatcgca agctcggcgt ctccggccgc cgcgagttgg tgaatgcgct
        59951  cccgggtcgt tgacgcccgc ggccgtcccg ttccggacga tcacaatcct
        60001  ctgtgcctgt cagccctgtg ggtgcggcgt cgttgtcgta cgcccgtccg
        60051  gtcatgacag caatcctcaa caaaagctca aacggaagcg ggggcaggtg
    25  60101  catccacccc atcaccaatg tctcgtatgt tacttttggc ccaccgtagt
        60151  tgggactctg cggaccgtac ctcgtcaagt acggaccacg gatttagggg
        60201  tgcgtagatg tgtcgcgttg acgtgccgag ttgtggctaa ctacgctccg
        60251  tttcacggaa atcgaaactc ttcggcgtga atcgcggcga acttcgatcg
        60301  aattctggat ccgtttctc agaaggctaa aagacgacgg gggttatccg
    30  60351  aaaggatggc ggtctcgtga ctatcactca cctcactgac tcgaatcacc
        60401  gtacgaccgg cggggtgata tccgctcaga ccgcaccggc cggggagtcc
        60451  gtcggccccg ggctgatggc gtccctcgac cgtgacctca ccatcaagca
        60501  cgccaaccag gagttccgcc gccgcttcga cgattccgcc ggtgacgtct
        60551  gcggtcgcag cttccgggac ctgatgcacc cgagcgtgca gcagccgctg
    35  60601  atgcgccagt tctcccggct gatcgagggc aagcggcacc gcttcgcctc
        60651  gcacgtggtc gccgtgggcg cccaggacgc ggccttcgcg ggcaccctca
        60701  ccgcctccgc ggtcaccggc aagacccegg acatcgccgg gatcctggtg
```

```
60751  ctcatggact cgtccggcgc cgccgacgcc gccgatgccg gcgtcgtgac
60801  cagccagaag aagttcctga ccgagatcga tgcgcgcatc ctcgaaggca
60851  tcgcggccgg gctgtcgacc attccgctcg cctcgcggct ctacctcagc
60901  cgccagggcg tggagtacca cgtcaccggg ctgctgcgga agttgcgggt
60951  gcccaaccgc gccgcgctcg tctcgcgcgc ctactccatg ggcatcctga
61001  acgtcggcac ctggcccccg aaggtcgtgg acgacttcat caagtgacgc
61051  ccggccggcc tgtcccgcac cggccggcca ccgtcggacc gctgtcccgc
61101  acgcccgtgt cgccgacgcc cgcctctcac ggggcgatgc ccggcggcac
61151  gggcgtttcg tgccctggtc gatgtttatc cattctcacg ttacgaaccc
61201  gcaatgctca tgcgcgcgat tccggccact tttcgttccg tcgtaaccaa
61251  actgcccgtc atatcgtttc cactgaccgt ttctgtggtc cgcggtgtgc
61301  cacctgggcc ccactccgac gcgccgcgca cgggtcgcga gcggcctccg
61351  aatactcacg gtaacccccg ttcctcctgg tggagttccc cgcgcgtcac
61401  attcctccgt gcctcgtgcg gcgtggctcg cccgctttcc ccgccgccgc
61451  gacagccatt gttacaacgg cacgacagag cggggtggt cgacccgtcc
61501  gcggtgcggc gcagtcaagc ggaacctctg ctcgatgttc gtgccacccc
61551  gaactctgtg ttcttggtga ccgttttcac caatgccccg cccgttcggt
61601  ggcgaatccg ccgcccgttc ctggccacaa cggatcgctt cggccgccca
61651  ccacggccgg cgcagccgag gcttgttgga atcaccccca cccagtcggg
61701  gcgccgttgt ggcccggaca cgcaggagga acaccgtgga tgctgagggc
61751  cgccgcaggg acatgctgga gctgatccgg cgcagcggat ccgccgatgt
61801  cgtgcgtctc gccgaggagt tcgccgtcag caaggaaacg gtccgccggg
61851  acctgaacgt cctggagggc catgggctga tacggcggcg gcacggtggc
61901  gcctatccca tggtgcggcc gggctccgag gccgtcttcg tctcccggac
61951  cgcacagccg atcccggagg agtcccggat cgccaccgcg gccgccgaac
62001  tcctcagcga ggccgagacg gtcttcatcg acgagggctt caccccgcaa
62051  ctcatcgccg acgccctgcc gcgcgaccgg ccgctgacca tcgtcaccgc
62101  ctcgctcccg gtggtcagcg ccttcgcgac gagcccacag gccaacgtgc
62151  tgctcctggg cggccgggtc cgccggggca cgacggccac cgtcgaccac
62201  tgggccgtcc acatgctgtc cggcttcgtc atcgacctgg ccttcctcgg
62251  cgcggagggg atctcgcgca ggtacggcct gaccacccc gacccggcgg
62301  tcgccgaggt caaggcccag gccatccgcg tcgcgcgccg cccggtcctc
62351  gccggggtgc acaccaagtt cggcacggcg agcttctgcc ggttcggaga
62401  ggtgggcgac ctggagacga tcgtcaccgg cgccggcctg cccgtcgccg
62451  aggcccaccg ctaccacctc atgggcccca aggttttacg ggtgtgacgc
62501  cgccggggcg tcccgccgcc cgcccacggc ctggcgcgcg ccggtgcggg
62551  tcagccgcgt cgcgcgtccg tcggttcccg ccagagtgcg agcggtgtca
```

```
       62601  cccccacccg gtagccctcc tcgtcgagga gatggccgcc gtgcttcatg

       62651  tgccagaggg cgaccgcgcg ttcggtgacg gtcaggtccg cggcccgttc

       62701  ggcgagtcgg acctcgaagc gctgggcgtc gtcgagggag cgcacccagg

       62751  ccaccaggtg caggttgtgc cggctgatga cgctcgcgca caggcggacc

    5  62801  tcgcgcatcc cggtgacccg gcgggtcacc tcgcgcagcc tcgcggccgg

       62851  gacctgcccc cagaagctca ccgtgaccgg ccactcggac agcgggcggg

       62901  cgacctcgca ccgggcgtgc aacatgtccg cggcgaagag ccgttggacg

       62951  cgacggcgca cggtgtcggg gccggcgccg cactgctcgg ccagcgcgcg

       63001  gtaggtggcc cggccgtcca cggacagcgc ggtgaccagc tgttggtcga

   10  63051  gctcgtcgag gacgaaggcg ggggtgtccg tgcggtgtct ggacgcgtcg

       63101  gcggcgagcc gggcgacctg gtgccggccg agggcccgca ggcgccaccg

       63151  gctgccctcg gtgtgcaccg ggccggccag gtgcgtccgg gcggcccgga

       63201  cgccgtccaa cgcggcgagg tcgtgggtca cccaacggga gagcatggcc

       63251  ggatcgcggg ccatgacgtt gagttggagg tcacggtccc cggtgacgtg

   15  63301  ggagagcgcc accacgtgcg gcgcggcggc caacgcccgc gcgacgtcga

       63351  gcagtcggcc gggagcgcag tcgatctcga cgaacgccag gcacccctgg

       63401  cccgacgcgg ccagcaccgg ggccggatgg cagctgatcc aggcggcgcc

       63451  ggtctcgacc agccggttcc agcgcctggc caccgtcacg gcgtccagtc

       63501  cgaggacgga gccgatccgg gtccagctgg cccgcggcgt gatctgcagc

   20  63551  gcgtgcacca gggcctgatc cacatggtcc agggaccggg gggtctgccc

       63601  ggaatcctgc gccacgggcc acctccttgc gtgtttccgg cggattcggg

       63651  ccgccggtcg gctcaacctt cagcctggac tcgggtacgg ccggaccgta

       63701  ccaggcaacc cccggagcaa caggagtggg attcggcatg tcggcagtgg

       63751  aacgcgccaa ggaaatgcaa cccgagttgg agcggctgcg ccggtcgctg

   25  63801  caccgggagc cggagttggg ccttgcgctg ccgcgcaccc aggagaaggt

       63851  gctcgcggcg ctcgacgggc tgccgttgga gatcaccccg ggcaccggcc

       63901  tcacctcggt gaccgccgtc ctgcgcggcg gccgccccgg cggcgcagtg

       63951  ctgctgcgcg cgacatgga cgcgctgccg gtcaccgagg agagcggtgt

       64001  gccgtacgcc tcggagatcc ccggccggat gcacgcctgc ggacacgacc

   30  64051  tgcacaccgc cggactcgtc ggcgcggcca ggctgttggc cgaacgccgg

       64101  gaacgcctgc acggcgacgt ggtgttcatg ttccagcccg gcgaggaggg

       64151  ccacaacggg ccggcgcga tgatcgagga gggcgtgctg gacgcggcgg

       64201  gcaagccgct cgacgcggcg tacgccctgc acgtggcgtc caaccaggtg

       64251  cccggcggcg tggtcatcac ccgctccggc accatcacct cggcctccga

   35  64301  cctgctgacg gtgacggtgc ggggcgaggg cggccacggc tcgacgcccg

       64351  ccacggccaa ggacccggtc ccggcggcct gcgagatggt caccgcggtg

       64401  cagaactggg tgacccgcgc cttcgacatc ttcgacccgg tcgtggtcac
```

```
        64451  cgtcggcacc  ttccacgccg  ggacgaaggc  gagtgtcatc  cccgacaacc

        64501  gccgagttcc  aggccacggt  gcgcagttat  tccgaatccg  cccgggaccg

        64551  gctcgccgaa  gggaactgcc  ccgtctggtg  cgggccatcg  cggccgccca

        64601  cggcctggaa  gcggaggtcg  actacctccg  gcactatccg  gtcaccgtca

    5   64651  acgacacgga  cgagaccgcc  ttcgcggtgg  ccaccgcccg  cgacctgttc

        64701  gcgcccggcg  aggtgtggga  gtcgccgata  cccaccaacg  gctccgagga

        64751  cttcgcgtac  gtcctccggc  gcgtccccgg  cgcgatgctg  ctggtgggcg

        64801  ccgcccccga  ggggagcgac  tggcagcacg  cgccgatgaa  ccactcgccg

        64851  cgcgccgtct  tcgacgaccg  ggtgctgtac  cggcaggcgg  cgctgctgac

   10   64901  cgaactcgcc  gcgcgccggc  tcgcggccac  cgcgccggcc  gagcccgccg

        64951  tggcgggatg  acaccgccgc  cgcacgcgtg  gggcccgcac  gcgtgcggcg

        65001  tacgggccct  ggccggtctc  cttcccggct  acttgcgcgc  ggggcactcc

        65051  ttccacgcga  agtggtagag  ggtgttgatg  ctgccgtcgg  tggagtcatg

        65101  gccatgtagc  tggtggcctt  ggggtcagcc  ccgctgccgt
```

   15

<u>SEO ID NO. 2</u>

<u>DNA sequence: Nys2</u>


```
   20       1   gaattcggat  cgatgctgct  cgtggtgtgt  gccgcgggag  gtgccgacgg

           51   tgccgtgcgt  gaatcggggc  gtggtcgccg  tcgcgatcgc  ggtggtgtcg

          101   gtgctgctcg  tgctgaccgt  ggtggtggcg  acggcctact  tgccgtgagc

          151   ccggtccttg  cgccgcgacc  tgggcagcgg  atgcacctct  tcggtcgacg

          201   ccaaggggcg  gtggcgcagg  tcgagatagg  cgcggacgga  gacgccgggc

   25     251   tgccccggcg  tctgccaggg  cgcggagggg  ccgctggaac  gccacagacg

          301   gcccagctcg  gccgcggcaa  gggtggcggt  ggcttcgtcg  gcggcggtga

          351   cgtcgatgac  gaccatgccg  ggttcgctga  catgttcctc  gtggatatgc

          401   acagttctgt  atgacctatt  tcgccccgtg  gcgtaaggag  tagccggcag

          451   gtttcatccg  aaggtggggc  gcgggagcgg  cgagttgacc  gtgaggtgcg

   30     501   tgcggtgttc  ccggtcgcgc  cggggggtgg  gggccggcag  gtggaccgcc

          551   gtggtcagca  gggcgggggt  gctgtgccag  cggccgggga  agaccgtgac

          601   ccggcgaccg  cccagcaggg  ggcccgggac  caggagtcgg  gcggtgaagg

          651   cgccctgtgt  cgggtcgatg  acgatctcgg  cctcggagaa  gtccagttcg

          701   cgctgggtca  ggggatacca  cgccttgaag  acgctctcct  tggcgctgaa

   35     751   gagcagccgg  tcccagtgga  cgtccggacg  atgggccgcc  agggccacga

          801   ggtgcggccg  ttccgagggc  agggcgatgg  cgttcaggac  gccggccggc

          851   agcggaccgt  tcggttcggc  gtcgatgctc  accgcggccg  acagctccgc
```

```
 901  gggggagacg gcggcggcgc ggtagccggc gcagtgcgtc atgctgccga
 951  cgatgcccgg cggccactgc ggggcgccgc gccgattggg cagtatggcc
1001  cggtccgggt ggccgagccg gcgcagggcc cggcgggcga gatggcggac
1051  ggtggtgaac tcgcgctgcc gggactccac ggcccgggcg atgacctcgc
1101  gttccgagga gagcagccgg tcgccggggc gcggccggtc gtcgtacgcc
1151  gcctcggtgg cgaccgtggc gggcaggatc agttcgatca ccgcatacct
1201  ccggcgaacg gtaagaacag ggggttgctt cggggcacgg ttccgtcctt
1251  gacggggcgc cgtgggcggc cgggtcagcc ggccgcggcg cccgcggtgg
1301  gagtgtgggt gcgggctgcg tgcagccggg cgtacaggcc ctgcgcgcac
1351  aggagttgat cgtgggtgcc ctgctcgacg atgcggccgg cgtccatcac
1401  gacgatcagg tccgcgtcgc ggatcgtgga caggcggtgc gcgatcacga
1451  agctggtccg gcccgcccgc agggagttca tggcgcgttg gatcaggacc
1501  tcggtccggg tgtccacgga gctggtggcc tcgtccagca cgaggacggc
1551  cggtctggcg aggaaggccc gggccacggt cagcagctgc ttctcgccgg
1601  cgctgacggt gcccgactcg tcgtccagca cggtgtcgta gccctgcggc
1651  agggtgcgga tgaagcggtc ggcacaggtc gcgcgggccg cctcctcgat
1701  gtccgcacgg caggcaccgg gtgcgccgta cgcgatgttc tccgcgatgg
1751  tgccgccgaa cagccaggtg tcctggagca ccagcccgaa gcgggaccgc
1801  aggtcgtcgc gggtcatcgt cgcggtgtcg gtgccgtcca ggaggatgcg
1851  gccggagtcc ggttcgtaga agcgcatcag gaggttgccg agggtggttt
1901  tgccggcgcc ggtggggccg acgatcgcca cggtgctgcc cggttccacg
1951  gtcagcgaga ggttctcgat gaggggcgtg tcggggagt agcggaagga
2001  cacgtcggtg aactcgacgc ggccctcggc gcgggcgggc gtgccgggcc
2051  ggagcgggtc cggggcctgc tcggggcgt cgagcagggt gaagacgcgc
2101  tgggcggagg cgatgccgga ctggagccgg cccgccaccg aggcgatctc
2151  cacgatcggc tggctgaact ggcgggcgta gaggatgaac gcctgcacgt
2201  caccgagggt cagggtgccg tttatgacct tccaggcgcc gatgacggcc
2251  accagcacat agccgaggtt ggcgacgaac atcatgaccg gttccatggc
2301  accggaggcg aactgcgcct tggccgcagc ccggtagacc gcgtcgttgc
2351  aggcgtcgaa gcgctcctcg gccgccgcgc gccggtcgaa gcccttgatc
2401  agcgcatgac cggtgcacac ctcctccaca tgggcgttga gggtgccgtt
2451  cgcggaccac tgcgcggcgt agtggggctg cgcgcgcttg ctgatccggg
2501  ccgcgatcaa cgccgagacc ggcacgctga gcagcatcac cacggccagc
2551  gacggcgaga tcaccagcat cagcaccagc atcgtcaaca gcgagaagat
2601  cgaggtgatc agctcggcga gggtctgctg gagggtctgt tggaggttgt
2651  cgatgtcgtt ggtggtgcgg ctgagcagct caccggccgg ctgccggtcg
2701  aagtggcgca gcggcagccg ggtcagcttc tcccgggcgt cgcggcgcag
```

```
       2751  ttcgtggatg  gtgcgccaca  ccgcggacgc  caccagccgg  ccctgcgcca
       2801  gcatgaacag  cgacgccacg  acgtagagcg  ccagcaggac  cagcagcagc
       2851  cggccgatcg  cggcgaagtc  gatcccgggc  gccgggcccg  ggacgccgcc
       2901  gagcacgccg  tcggcgatca  gatcggtgac  ccggccgagc  agcagcgggc
  5    2951  cgaacgcgtt  gagcacgatc  ccgccgacgc  ccatcgacac  ggcgagtgcc
       3001  acggagcggc  ggtgcggacg  cagcaggccg  acgagccgac  gtgccggccg
       3051  gggcgcggtc  cgctcctcct  caaggtcgtc  cggcgaggcc  atgggcggct
       3101  tcctcctcgg  tcagctgcga  gagcgcgatc  tcgcggtagg  tcgggctggt
       3151  gcgcagcagc  acgtcgtggg  tgccctgggc  gaccacgcgc  ccgcggtcca
  10   3201  gcaccacgat  ccggtccgcg  tcgcggccgg  cggagatccg  ctgcgccacg
       3251  gtgatcaccg  tggcgcccgc  ggtgtacggc  accagcgcgg  tccgcagcgc
       3301  cgcatcggtg  gcctggtcga  gcgccgagaa  acagtcgtcg  aagagataga
       3351  tctccggccg  gcgcagcaac  gcccgggcga  gggacaggcg  ttggcgctgg
       3401  ccgccggaga  cattgccgcc  gccctgggtg  atctccgcgt  cgaggccgtc
  15   3451  cggcatccgc  gccacgaagt  cggccgcctg  ggcgacccgc  agcgcctccc
       3501  acagctcctc  gtcggtggcg  tccggccgcc  cgaagcgcag  attgctcgcc
      .3551  acggtgccgg  agaacaggta  cggccgctgc  ggcacgaacc  cgacggcggc
       3601  ggcgagcgtg  gccgcggtca  gctcgcggac  gtcggtgccg  ccgacccgca
       3651  ccgcgccctc  ggtggcgtcg  gccagccgca  gcaccaagtt  caacagggtc
  20   3701  gtcttgccgc  tgccggtgct  gccgagcacg  gcgatccgct  cgcccggctc
       3751  gacggtcagg  tcgacgtccc  gcagcacggg  ctcctcggcg  cccgggtagc
       3801  ggtacccggc  cgcgcacagt  tcgatccggc  cggcgggccc  gcgcaccggc
       3851  tgcggcgcgg  ccggcggcgc  cacgctcgac  ccggtgtcca  ggacctccgc
       3901  gatccggccg  gcacagaccc  gggcccgcgg  caccgacagg  aacacgaagg
  25   3951  cgagcatcac  gacggacatc  aggatcagcg  agagatagct  caggagggcg
       4001  ctgagcgagc  cgatcggcat  ccggcccgcg  tcgatccggt  gggagccggt
       4051  ccacagcagg  gctacggtga  aaccgttcat  cagcagcagc  acgaccggca
       4101  gcatcgtcgc  gatcagccga  cccacccgcc  gcgacaccac  gaggaacgcg
       4151  tcgttggtct  gcgcgaaccg  cgcgcgctcg  tggtcgtcgc  ggacgaagga
  30   4201  ccggaccacc  cgcaccccgg  tgatcgcctc  gcgcagcagc  cgccccagcc
       4251  ggtccagggt  cagctgcatc  cgcgcgtaca  gggtgcccat  ccgggccagc
       4301  agcaggccga  agcagaccgc  caccaccagc  accagcgcca  ccagcagcag
       4351  tgccagcgga  acgtcctggc  gcagcgccag  cagcacgctg  cccaggcaca
       4401  tcagcggcgc  gcagacgacg  atgccgaagc  cggtctgggc  gaggttctgc
  35   4451  acctgctgca  cgtcgttcac  cgaccgggtc  agcagggagg  gggtgccgaa
       4501  ccggccgatc  tcgcggcggg  agaagtccag  gatgcggcgg  aagagcgcgg
       4551  accgcagatc  gcggcccatc  gccgtcgcgg  tccgggcggc  cagcgcggcc
```

```
        4601  gcaccgagcg  cggcggcgat  ctgcaccagc  gccaccacgc  ccatcaccac
        4651  acccagctcg  gtgatgcgcc  cgccgtcgcc  gcgcaccacc  ccctggtcga
        4701  taagtgcggc  gcccagcgtc  ggcagcagca  aagtgcccag  gatctggacg
        4751  agttgaaggg  cgacgagagc  ggcggtggcc  caggcgtagg  ggcgcagctg
  5     4801  tgcccgcaga  agtctcaaca  gcacggagga  acaccccggt  tgacggcggc
        4851  gtgcggccgc  gcggtggacg  gagtcggccg  gcccgccgac  ccggtcattg
        4901  caccgcagtc  cgctccgaaa  tttcactagt  gttgggggtg  gcaacggact
        4951  tttgacggcc  cggtactcgt  gaattcccta  gaaagcccgg  gatgcgttga
        5001  cagcatcttc  cgcggcttgc  gagcgtgcgg  gtgtctatcg  tccggactgc
  10    5051  gtattcgacg  ccaggtcgtg  gccgagctca  agcgtttgga  cggtctcttt
        5101  cgtgttcgag  aagggtttgc  catgtccaaa  cgagcgctga  tcaccggaat
        5151  caccggccag  gacggctcct  atctcgcgga  gcacctgctg  tcccagggct
        5201  accaggtgtg  gggtctgatc  cgcggccagg  ccaatccccg  caagtcccgg
        5251  gtcagccgcc  tcgcctccga  actcgacttc  atcgacgggg  acctgatgga
  15    5301  ccagggcagc  ctggtctccg  ccgtcgacac  cgtgcagccc  gacgaggtct
        5351  acaacctcgg  cgccatctcg  ttcgtgccga  tgtcctggca  gcaggccgag
        5401  ctggtcaccg  aggtcaacgg  catgggcgtg  ctgcgcatgc  tggaagccat
        5451  ccgcatggtc  agcggactgt  ccacctcccg  cacggtcagc  ccgcgcggcc
        5501  agatccgctt  ctaccaggcg  tccagctcgg  agatgttcgg  caaggccgcc
  20    5551  gagacgccgc  agcgcgagac  caccctcttc  cacccgcgca  gcccctacgg
        5601  cgcggcaaag  gcgtacgggc  actacatcac  ccgcaactac  cgcgagtcct
        5651  tcggcatgta  cgcggtctcc  ggcatgctct  caaccacga  atccccgcgc
        5701  cgcggccagg  aattcgtcac  ccgcaagatc  agcctggcgg  tcgcccgcat
        5751  caagcaaggc  ctccaggaca  agctggcact  cggcaacctc  gacgcggtgc
  25    5801  gcgactgggg  ctatgccggc  gactacgtcc  gcgccatgca  cctgatgctc
        5851  cagcaggacg  ccggcgacga  ctacgtcatc  ggcaccgggc  agatgcactc
        5901  ggtgcgcgac  gcggttcgga  tcgcgttcga  acacgtcggc  ctgaactggg
        5951  aggactacgt  cgtcatcgac  cccgacctgg  tgcggccgc  cgaggtcgag
        6001  gtgctgtgcg  ccgacagcgc  caaggcccag  gaccgcctcg  gctggaagcc
  30    6051  ggacgtcgac  ttccccaccc  tcatgcgcat  gatggtcgat  tccgacctgg
        6101  cgcaggtttc  ccgcgaaaac  caatacggcg  acgtgctgct  cgccgccaac
        6151  tggtagcagt  tctcaagctt  tcgaaaacta  gtgaattcct  gccggaattc
        6201  cgacgacact  cgccacatgg  attcccccgg  tgagtggcga  atccaggtgg
        6251  cgaatccgaa  cgtaccgccg  aaacggcgtg  gagaagtcgg  acgccattca
  35    6301  cgtgcggccg  tcggctgtcg  agatgggttg  agttgagatg  gacaacgaac
        6351  aaaaactccg  ggattacctc  aagcttgcga  cggccgacct  tcgacgcacc
        6401  cggcggcgcg  tccacaagct  ggagtcggcg  gcccaggaac  cggtggccat
```

SUBSTITUTE SHEET (RULE 26)

	6451	catcggcatg	acctgtcgct	accccggcgg	cgtccgcagc	cccgaagacc
	6501	tctggcgcat	ggtcgaggcc	ggcgagcacg	gcgtcacccc	gttccccacc
	6551	gaccgcggtt	gggacctgga	ggcgctggcc	gccgcgccga	ccgcctccgg
	6601	cggattcctg	cacgacgcac	ccgacttcga	cgcggacttc	ttcggcatct
5	6651	cgccgcgcga	ggcggtcgcc	atggacccgc	aacagcgcgt	cgtcctggaa
	6701	tccgcctggg	aggcgttcga	acgcgccggc	atcgacccga	cgtccgtgaa
	6751	gggcagccgc	accggagtct	tcatcggcgc	gatggcccag	gactaccggg
	6801	tcggccccgc	cgacggcgcc	gagggcttcc	aactcaccgg	caacaccggc
	6851	agcgtgctgt	ccggccgcat	ctcctacacc	ttcggcacgg	tcggcccccgc
10	6901	cgtcaccgtc	gacaccgcct	gctcctcctc	cctcgtcgcc	gtccacctcg
	6951	ccacccaggc	gctgcgggcc	ggcgagtgca	ccctcgccct	cgccggcggc
	7001	gtcaccatca	tgtccggccc	cggcaccttc	atcgaaatgg	ccgccaggg
	7051	cgggctctcc	gccgacggcc	gctgccgctc	cttcggcgac	accgccgacg
	7101	gcaccggctg	ggccgaaggc	gtcggcatcc	tcgtcctgga	acggctgtcc
15	7151	gacgccgtcc	gcaacggcca	cgagatcctc	gccgtcgtcc	gcggcaccgc
	7201	cgtcaaccag	gacggcgcct	ccaacggcct	gaccgccccc	aacggcccct
	7251	cccagcagca	ggtcatccag	caggccctgg	tcaacgcccg	actcgccgcc
	7301	ggggacatcg	acgtcgtcga	ggcgcacggc	accggcacca	ccctcggcga
	7351	ccccgtcgag	gcccaggccc	tgctcgccac	ctacgggcag	aaccgcccgg
20	7401	cggaccggcc	gctgctgctg	ggctcggtca	agtccaacct	cagccacacc
	7451	caggccgccg	ccggcgtcgc	cggcgtgatc	aagatggtca	tggcgatgcg
	7501	gcacggcacc	ctgccgcgca	ccctgcacgc	cgaggagccc	acccaccacg
	7551	tcgactggtc	gcagggcgcc	gtgcggctgc	tgaccgacac	caccgactgg
	7601	cccgccaccg	gggcgccgcg	ccgcgccgcc	gtctcctcct	tcggcatcag
25	7651	cggcaccaac	gcccacacca	tcatcgagca	ggcccccgaa	ccgcagcccg
	7701	aggacgccgc	gaccgcgcag	gacgacgccg	ccggcagcac	gccggccacc
	7751	gcccccgtag	tgcccggcgt	cgtaccggtc	ctgctctccg	gccgcacccc
	7801	ggacgccctg	cgcggccagg	ccgcggccct	gcgcgccgcc	ctcgacaccg
	7851	gccggcggcc	cgacctgctc	gacctcgcac	actccctcgc	caccacccgc
30	7901	gccgggttcg	agcaccgcgc	cgtcctcctc	gccaccgacc	accccgccct
	7951	gaccgacggc	ctcaccgccc	tcgccgacgc	cgacgacccg	gccgccgccc
	8001	ccgcctggat	caccggcacc	acccgggccg	agacccggct	cgccgtcctg
	8051	ttcaccggcc	agggcgccca	acgcctcggc	gcgggacggg	aactcgccgc
	8101	ccgtttcccg	gcgttcgcca	ccgccctcga	cgcggcgctc	gacgccttca
35	8151	ccccgcacct	cgaccgcccc	ctgcgcgagg	tcctgtgggg	caccgacgcc
	8201	gccctgctcg	accgcaccgc	atacgcccag	ccggccctct	cgccgtcga
	8251	agtggcgctc	taccggctga	tcgaatcgtt	cggcgtccgc	cccgaccacc

```
        8301  tcgccggcca ctccgtcggc gagatcgtcg ccgcgcacct cgccggggtc
        8351  ctctccctgg ccgacgccgc caccctcgtc gccgcccgcg gtcgcctgat
        8401  gcaggcgctg cccgacggcg gggcgatgat cgccgtccag gcgtcggaag
        8451  ccgacgtcgc cccgctgctc gccgggcacg aggaccaggt cgcgatcgcc
   5    8501  gccgtcaacg gcccctccgc cgtcgtcctg tccggcgccg aagccaccgt
        8551  caccgcgctc gccgaacagc tcgccgccga cggccgcaag acccgccggc
        8601  tgcgcgtctc gcacgccttc cactcgccgc tcatggagcc gatgctcgac
        8651  gccttccgcg ccgtcgtcga agacctcacg ctccagccgc cgctcctgcc
        8701  ggtcgtctcc aacctgaccg gcaagcccgc caccgtcgcc caactcacct
  10    8751  ccgccgacta ctgggtcgac cacgtccggc acgccgtccg cttcgccgac
        8801  ggcatcgact ggctcgcccg gcacgacacc accgccttcc tcgaactcgg
        8851  ccccgacggc gtgctgtccg ccatggccca ggactgcctg gacgccgccg
        8901  acgcagacgc cgtcaccctc cccgccctgc gcgccgggcg ccccgaggag
        8951  cacaccctca ccaccgccct cgccggtctg cacgtccacg gcgccaccct
  15    9001  ggactggacc ggctgcttcg ccggcaccgg cgcccgccgc accgacctgc
        9051  cgacctacgc cttccagcgc cgccgctact ggcccaaggc cctccagagc
        9101  ggcaccgccg acctgcgctc ggtcggcctc ggtgccgccc accacccgct
        9151  gctctccgcc gccgtctccc tcgccgacgc aggcggcacc ctgctcaccg
        9201  gccgcctctc ccggcagacc cacccctggc tcgccgacca caccgtccgc
  20    9251  ggcaccaccc tgctgcccgg taccgccttc ctcgaactcg ccgtccgcgc
        9301  cggcgacgag gtcggctgcg accgcgtcga ggaactcacc ctcgccgcac
        9351  cgctcctgct gcccgaacag ggcggcgtcc aggtccagtt gtggatcggc
        9401  aaccccgacg tgtccggtcg ccgcaccgtc aacgtccacg cccgcccccga
        9451  caccggcgac gacacccccct ggaccgccca cgccaccggc gtcctcacca
  25    9501  ccgccgacgc ctcccgccag ctcccggctt cgtccgagca gggcggcacc
        9551  ccccctcgccg gcgaccccca ccccgccctc gacgcggccc agtggccccc
        9601  ggccggcgcc gaaccgctgc cgctggacgg ccactacgac cgcctcgccg
        9651  acggcggctt cggctacggc ccggtcttcc agggcctgcg cgccgcctgg
        9701  cgcggcggcg acgtcgtcta cgccgaggtc gagctgcccg aggccggccg
  30    9751  gtccgacgcc gaggcgttcg gcctccaccc cgccctgctc gacgccgccc
        9801  tgcacgccgc gcccttcacc ggcctcggcg aacgcggccg gggcggcctg
        9851  ccgttctcct gggagggcgt ctccctccac gccggcggcg ccaccaccct
        9901  ccgcgtccgc ctgaccccgg tcgccgacga cgcgctcgcc ctgaccgtcg
        9951  ccgacggcac cggcgcgccc gtgctgtccg tcgactcgct cgtcctgcgc
  35   10001  agcgtggcga cccaacagct cgacacggcc gccgccgtcg cccgtgacgc
       10051  cctcttccgc ctcgactgga ccccccgtcca gccgaccgcc accgaccccg
       10101  ggcccgtcgc cctcctcggc gccgacccct cggcctgct cacccacgcc
```

```
        10151   ggattcgccg  acgccccggc  atacccggac  ctcgccgccc  tcgccgcggc

        10201   ggacggcccg  gtcccgacca  ccgtcgtgct  gtccctcgcc  ggcaccgggg

        10251   acgacgcggc  cgacccggcc  cggtccgcac  accgctgcgc  cgcggaggcc

        10301   ctcgccgccg  tacagacctg  gctcgaccac  catgagcgct  tcgccgccgc

  5     10351   ccgcctggtc  ttcgtgaccc  gtggtgcgac  ggtcgggcgt  gatgttgctg

        10401   ctgctgcggt  gtggggtctg  gtgcgttcgg  cgcagtcgga  gaatccgggg

        10451   tgttttgctc  tggtcgatct  ggatccggat  ggtgcggtgg  gtgcggctgc

        10501   gctcgtcgct  gcgttggtca  gtggtgagcc  gcagcttgcg  gtgcgcggtg

        10551   atgtgttgcg  ggtcgcgcgt  ctggtgcggc  ggccgctcac  cgaggtcggt

  10    10601   gcgggtgctg  atggcaccgg  ggatggcgtc  ggggatggct  ctggtgtgtc

        10651   gttctcgggt  gagggtgcgg  tcctggtcac  tggtggtacg  ggtggtctgg

        10701   gtgcggtgtt  ggcgcgtcat  ctggtggccg  agtatggggt  gcgggatctg

        10751   ctgttggtca  gtcgcagtgg  tgaacgtgcc  gtgggtgctg  gggagttggt

        10801   ggcggagctt  gcgggtgtgg  gtgcgcgggt  gcgggtggtt  gcgtgtgatg

  15    10851   tgaccgatcg  tgccgcggtg  gtggagttgg  ttggcgggca  tgcggtgtcc

        10901   gcggtggttc  atgcggctgg  tgtgctggat  gacggcatgg  tgggtgcgtt

        10951   gaccggggag  cggttgtccg  cggtgctgcg  gccgaaggtg  gatgctgtct

        11001   ggcatctaca  tgaggcgacc  cgcggcctgg  acctggacgc  gttcgtcgtc

        11051   ttctcctccc  tcgccggggt  cttcggcagt  cccggccagg  ccaactacgc

  20    11101   ggccgcgaac  gccttcctgg  acgcgctgat  gacgcggcgc  cgggcggagg

        11151   gactgcccgg  cctgtcactc  gcatggggac  cgtgggagca  gtcgggcgga

        11201   atgacgggca  ccctgacgga  cgtcgacgcc  gaacggctgg  cccgctccgg

        11251   tgtcccgccg  ctctccgtgg  cgcagggcct  ggccctcttc  gacgctgccg

        11301   tggccgggac  cgacgccacc  tgcgttccgg  tccgcctgga  cctccccgtc

  25    11351   ctccgcgcac  ggggtgaagt  gccgccgctg  ctgaggtcgt  tgatccgcgt

        11401   ccgggcgcgc  cgagccgccg  tcgccggctc  cgccaccgcg  ggcaacctcg

        11451   cccagcgcct  gcgccgcctg  gacgaggacg  gccgcgacga  gatggtcctg

        11501   gacctcgtcc  gcggtcaggt  cgccctcgtc  ctcggccacg  cgaccggtgg

        11551   cgacgtcgac  gccggccgtg  ccttccgcga  cctcggcttc  gactcgctga

  30    11601   ccgccgtcga  actgcgcaac  cgcctcaaca  ccgtcaccgg  cctgcgcctg

        11651   cccgccaccc  tggtcttcga  ctacccgacc  gtccggcacc  tcgccacgta

        11701   cgtcctggac  gagttgttgg  gcacggatgc  cgaggtggcg  accgtgcagc

        11751   cggccgcggt  tgcggtggcg  gacgatccga  tcgtcatcgt  gggcatggcc

        11801   tgccgctacc  ccggtggcgt  cagctccccc  gaggacctgt  ggcgcgtgct

  35    11851   caccgaaggc  accgacgccg  tctcgggctt  cccgaccaac  cgtggttggg

        11901   acgtcgaatc  cctctatcac  ccggaccctg  accaccgggg  tacctcctac

        11951   acgcgctcgg  gtgggttcct  gcatgaggcg  ggggagttcg  atccgggggtt
```

```
12001  cttcgggatg agtccgcggg aggcgttggc gaccgattcc cagcagcggt
12051  tgttgttgga gtcgtcgtgg gaggcgatcg agcgggccgg tattgatccg
12101  gtgagtttgc ggggtagtcg dacgggtgtg ttcgcggggg tgatgtacag
12151  cgattacagc gcgatgttgg cgagtccgga gttcgagggt ttccagggca
12201  gtgggagttc gccgagtttg gcgtcgggtc gggtggccta cacgttgggg
12251  ttggaaggcc cggcggtgac ggtggatacg gcgtgttcgt cgtcgttggt
12301  ggcgatgcac tgggcgatgc aggcgttgcg tagtggtgag tgtgggttgg
12351  cgttggccgg tggtgtgacg gtgatgtcga cgcctgcggt gtttgtggac
12401  tttgctcggc agcggggttt gtcgccggat ggccggtgca aggcgtttgc
12451  ggatgcggcc gatggtgtgg gctggtccga gggcgtcggc gtgttggtcc
12501  tggagcggca gtcggacgcg gtgcgcaatg gtcacgagat tttggctgtg
12551  gtgcggggtt cggcggtcaa ccaggatggt gcgtccaatg gtttgacggc
12601  gcccaatggt ccgtcgcagc agcgggtgat ccggcaggcg ttggccagtg
12651  gtggtctgac ggccggtgac gtggatgtgg tggaggcgca tggtacgggt
12701  acgacgctcg gtgatccgat cgaggcgcag gcgttgttgg cgacgtatgg
12751  gcgggatcgt gagcctgagc ggccgttgtt gttgggttcg gtgaagtcga
12801  atctgggca tacgcaggct gctgcgggtg tggcgggcgt tatcaagatg
12851  gtgttggcga tgcggcatgg tgtggtgccg cggacgttgc atgtggatgc
12901  gccttcttcg catgtggact ggtccgaggg tgcggtggag ctgctcagtg
12951  agcaggcggc ctggccggag acgggtcggg tgcggcgggc gggtgtctcc
13001  tccttcggca tcagcggtac caatgtgcat gtcatcgtcg agcaggcgcc
13051  gggcgccaag gcgatcgccg cggccggtgc ggcgcggcgc acgccgggtg
13101  ccgtgccggt gctgctctcg gggcgtggcc ggagtgccct gcggggccag
13151  gccgcccgcc tgctcggaca cctccaggcc cgacccgacg ccgaactcgt
13201  cgatgtcgca ctgtcgttgg cgaccacccg gtcccgcttc gagcagcggg
13251  ccgccgtcgt ggcgcaggac cgcgaccagc tgatcgcctc gctggggggcg
13301  ctggccgccg accgccccga ccccgccgtc gtcgagggcg aggccgccgg
13351  acgcggccgg accgcggtgc tgttcaccgg acagggcagc cagcgggccg
13401  ccatggggcg tgaactccac gaggtgcagc cggagttcgc cgcggcgttc
13451  gacgcggtgt gtgccgtttt cgaccgcctg ttggaccggc cgctgcgcga
13501  ggtggtgttc gccgaggacg gcagcgacga ggccgcactg ctggacgaga
13551  ccggttggac gcagccggct ctgttcgccg tcgaggtggc gctgttccgg
13601  ctggtggaga gttggggtgt ccgtccggac ttcgtggccg gccattccat
13651  cggtgagatc gcggcggcgc acgtcgccgg ggtgctgacg ttggaggacg
13701  cctgccgtct ggtggccgcg cgggcgacgc tgatgcaggc gctgccgacc
13751  ggcggcgcga tgatcgcgat ccaggccacc gaggacgaga tcgcggcgca
13801  cctcgacgac acggtggcga tcgccgccgt caacgggccg cagtccgtgg
```

```
      13851  tgatctccgg tgacgaggag gccgccgaaa cgatcgccgc cacgttcgcc
      13901  gaacgcgggc gcaagaccaa gcggctgcgg gtgagccatg ccttccactc
      13951  gccgcggatg gacgggatgc tggacgcctt ccggatcgtc gccgaggggc
      14001  tgacctaccg ggcgccgcgc atcccgctcg tctccgacct caccggccgg
  5   14051  cgcgccgacg atgcggaggt gtgcaccgcg gagtactggg tgcggcacgt
      14101  ccgagaggcc gtgcggttcg ccgactgcgt gcggacgctg cgcgacgccg
      14151  gggccaccac cttcctggaa ctgggctccg acggcctgct gaccgcgatg
      14201  gccgaggaca ccctcggtga cgaccacgac gccgaactgg tgccgatgct
      14251  gcgcgccggg cgcgccgagg aactggccgc ggccaccgcc ctggcccgcc
  10  14301  tccaggtgcg cggcgtggac gtggactggg cggcgtacct cgccggcacc
      14351  ggcgcccgac gcaccgacct gccgacctac gccttccagc acgcgtacta
      14401  ctggccgcag ctgccgaccc cggccgccgc cctcgccgcc gccgatcccg
      14451  ccgaccagca gctgtgggcc gctgtggagc gcggcgacgc ccgcgaactc
      14501  gccgacatcc tcggcctggg cgaacaggac ctcacgccgc tggactccct
  15  14551  gctgcccgcc ctcacctcgt ggcggcgcgg caaccaggag aagcacctcc
      14601  tggacaccct gcgctaccgc gtggagtgga cacgactgag caagccgacc
      14651  gccccggtcc tcgacggcac ctggctgctg gtcgcctccg acgccaccgc
      14701  ggccgaccag ccagccctcc tcgacggcct ggccgacgcc ctcggctcgc
      14751  acggcgcgcg ggtgcgtcgc ctgcttctgg acgactcctg cgcggaccgc
  20  14801  gcggtgctcg ccgaacgact ggcgcggacc gccgacgtgg acgccgcgac
      14851  ccaggtgctg tccgtgctgc cgctcgacga gcgggacgcc gacgactgcc
      14901  cgccgctcac ccgcggactg gcgctgaccg tcgcgctcgt ccaggccctc
      14951  gccgacaccg gcgcccaggg ccggctgtgg accgccaccc gcggcgccgt
      15001  ctccaccaac cccgccgacc cggtcaccca ccccgtccag gccgctgcct
  25  15051  ggggcctggg ccggggcgtc gccctggagc acccacggct gtggggcggc
      15101  ctggtcgacc tgccgcaggt cttcgacgag cgggccggac agcggctcgc
      15151  cgggatcctc gccgtcaagg acgcaccgga cggcgaggac caggtggcgc
      15201  tgcgggccac cggagtctcc ggccgccggc tcgtccgcca caccgtcgaa
      15251  gcgctgccca cggccgcgga gttcaccgcc accggcactg tcctgatcac
  30  15301  tggtggcacc ggtggcctgg gcgccgaggt cgcccggtgg ctggcccgcg
      15351  ccggcgccca gcacctcgtc ctgaccagcc gccgcggccc ggacgcgccg
      15401  ggcgccgccg aactccgggc cgaactggag ggctacgggc cgtcggtgtc
      15451  cgtcgtcgcc tgcgacgtcg ccgaccggga cgcgctcgcc gccgtcctca
      15501  ccgcactgcc cgaggaactg ccgctgaccg tgtcgtgca caccgcaggc
  35  15551  gtcggccact acggcccgct ggacaccctg agcaccgccg agttcgccgg
      15601  cctcaccgcc gccaagctcg ccggcgccgc ccacctcgac gccctgctcg
      15651  ccgaccgcga actggacttc ttcgtcctct tcggctccat cgccggtgtc
```

```
15701  tggggcagtg gcaaccagag cgcctacggc gccgccaacg cctacctcga
15751  cgcgctcgcc ctgcaccgcc gcgcccgcgg cctcgccgcg acctccgtcg
15801  cctggggccc gtgggccgag gccggcatgg ccgccgacga tgccgtttcc
15851  gagaccctgc gccgccaggg cctcggcctg ctcgacccgg ccccggccat
15901  gaccgagctg cgccgcgccg tcgtccggca ggacgtcacc gtcaccgtcg
15951  ccgacgtgga ctggcagcgc tacgcaccgc tgttcacctc cgcccggccc
16001  agcgccctga tcgccggcct gcccgaggtc cgcgccctcg ccgccgacga
16051  gcgcaccgag caggacgcca ccggcgcctc cgaggtcgtc acccgcgtcc
16101  gcgccctggc cgaacccgag caactgcgcc tgctgaccga cctcgtccgc
16151  accgagtccg ccaccgtcct cggccacagc tccgccgacg ccgtgcccga
16201  gggccgcgcc ttccgcgacg tcggcttcga ctcgctgacc gcggtcgagc
16251  tccgcaagcg cctgggcgcc gcgaccgggc tgtccctgcc cagcaccatg
16301  gtcttcgact acccgacacc gctggaactc gcccagtacc tgcgggcgga
16351  gatcctcggc gcggtgctgg aagtcgccgg cccggtcgcc accggcggcg
16401  ccgacgacga gccgatcgcc atcatcggca tggcctgccg cttccccggc
16451  ggcgtcagct ccccggaaca gctgtgggac ctggtcgcct ccggcaccga
16501  cgcgatcagc gagttccccg tcaaccgcgg ctggcagacc gggcacctct
16551  tcgacccgga ccccgaccgg cccggcacca cctactccac ccagggcggc
16601  ttcctccacg aggccgacga gttcgacccc accttcttcg gcatctcgcc
16651  ccgcgaggcg ctggtcatgg acccgcagca gcggctcctg ctggagacca
16701  cctgggagtc cttcgagcgc gccgggatcc gcccggaaac cctccgatcc
16751  accctgaccg gcaccttcgt cggctccagc taccaggagt acggcctggg
16801  cgcgggcgac ggcaccgagg gccacatggt caccggcagc agccccagtg
16851  tgctctccgg ccgactgtcg tacgtcttcg gtctggaagg cccggcggtc
16901  accgtcgaca ccgcctgctc gtcctcgctc gtggcgctgc acctggcctg
16951  ccagtcgctg cgcaacggcg agagcaacct ggccgtcgcc ggcggcgcca
17001  cgatcatgac gacgcccaac ccgttcatcg cgttcagccg gcagcgcgcc
17051  ctcgccaagg acggccgctg caaggcgttc tccgacgacg cggacggcat
17101  gacgctcgcc gagggcgtcg gcgtcgtcct cgtcgagcgg ctctccgacg
17151  cgcagcgcaa cggccacccg gtcctggccg tcctccgcgg ctccgccatc
17201  aaccaggacg gcgcctccaa cggcctgacc gcgcccaacg gcccgtccca
17251  gcagagggtc atccgccagg ccctcgccaa cgcccgcctc gcgcccgggg
17301  acatcgacgc cctggaggcg cacggcaccg gcacaccgct cggcgacccc
17351  atcgaggccc aggcactgtt cgccacctac ggccgcgacc gcgacccega
17401  gagcgcgctg ctgctcggct cggtgaagtc caacatcggc cacacccagt
17451  ccgccgcggg catcgccagc gtcatcaaga tggtcatggc gctgcgccac
17501  tccgaactgc cgccgaccct gcacgccgac gcgccgtcct cgcacgtgga
```

46

```
        17551  ctggtcggcc gggacggtcc ggctgctgac ccaggcgcgc gcctggccgg
        17601  agaccggtcg cccgcgccgg gccgcggtgt cctcgttcgg catcagcggt
        17651  accaacgccc atgtcctgct ggagcaggcg cccgtcgcgg acaccccggc
        17701  cgaggagcgg cccgccgtgg cgccggtccc gatcgccgcc ggcgtcgtcc
    5   17751  cgtgggtggt caccgcccgc agcgccgccg ccctgcgcgg ccaggccgag
        17801  cgcctcctcg cgcacgccga aaccgtcgga accgccctgc cggccgccgg
        17851  accgctcgac atcggcctgt cgctggtctc cgcgcgcgcc cgtttcgagc
        17901  accgtgccgt cgtcgtcccg cccgcgggca ccgacccgct ggccgcgctg
        17951  cgcgccgtcg cgacggacgg gccctcgccc gtggtcgccc gtggcgtggc
   10   18001  ggacgtcgag ggtcggacgg tgttcgtgtt ccccggtcag ggttcgcagt
        18051  gggtggggat ggggtcccaa ctccttgatg agtcggcggt gttcgcggag
        18101  cggattgccg agtgtgcggc ggcactcgcc gagttcaccg actggtcgct
        18151  ggtcgatgtg ctgcggggtg tggtgggtgc gccgtcgttg gagcgggtcg
        18201  atgtggtgca gccggcgtcg ttcgcggtga tggtgtcgtt ggctgcgttg
   15   18251  tggcgttccc gtggtgtgtt gccggatgcg gtggtggggc attcgcaggg
        18301  tgagatcgct gctgcggtgg tgtcgggtgc gctgtcgttg cgggacgggg
        18351  cgcgggtggt ggcgctgcgg agtcaggcca ttggtcgtgc gttggcgggg
        18401  cggggaggga tgatgtccgt cgcgctgtcg gtggacgtgc tcgaaccgcg
        18451  gttggtcgag ttcgaggggc gggtgtcggt ggccgccgtc aacggcccgc
   20   18501  gctccgtcgt ggtcgccggc gagcccgagg cgctggacgc gctgcacgcc
        18551  cggctgaccg ccgacgacat ccgggcccgc cggatcgcgg tggactacgc
        18601  ctcgcactcg caccaggtcg aggacctgca cgaggaactg ctggaggtgc
        18651  tggcggagct ggcgccgcgc acgtcggagg tgccgttctt ctcgaccgtg
        18701  accggcgact ggctggacac cgcgcggatg gacgccggct actggttccg
   25   18751  caacctgcgc ggacgggtgc ggttcgcgga cgcggtggcg gacctgctgg
        18801  cggcggagta ccgcgcattc gtcgaggtca gctcgcaccc ggtgctgtcg
        18851  atggcggtgc aggaggcgat cgacgaggcc ggcgtgccgg ccgtcgccgc
        18901  cggcaccctg cgccgcgacc agggcggcac cgaccgcttc ctgctgtcgg
        18951  ccgccgaggt cttcgtgcgc ggtgtggacg tggactgggc ggggctgttc
   30   19001  gagggaccg gtgcgtccg gatcgacctg cccacctacg ccttccaaca
        19051  cgaacacctg tgggccgtcc cgcccgcccc ggaggccgtc gccgccgccg
        19101  acccggacga cgcggccttc tggaccgcgg tcgaggacgg tgacgtctcc
        19151  gcgctgaccg ccgcgctcgg caccgacgag gactccgtcg ccgccgtgct
        19201  gcccgccctg acctcctggc gccgggcccg ccgcgaccgc tccaccgtgg
   35   19251  acgcctggcg ctaccgcgtc gcctggaaac ccctcggcgg caccctgccg
        19301  cacccgtccc tgaccggcac ctggctgctg gtcaccgccg acggcatcga
        19351  cgacaccgat gtggcagggg cgttggagac ctacggcgcc gaggtgcgcc
```

```
        19401  ggctggtcct  ggacgaggag  tgcgtcgacc  gcgccgtcct  gcgggagcgg
        19451  ctggccggcg  cggaggacgt  gaccggcatc  gtctccgtcc  tcgccgccgc
        19501  cgagcggacc  gacgcggtac  cgggcacctc  cctggtgctc  ggcaccgccc
        19551  tgaccgtggc  actgatccag  gccctcggcg  acgccgaaat  cgacgctccc
   5    19601  gtatgggcgt  tgacccgcgg  cgcggtctcc  accggccggg  ccgacgagct
        19651  gaccgcgccc  gtccaggcac  aggtcaccgg  catcggctgg  accgcggcgc
        19701  tggagcaccc  gcagcgctgg  ggcggcaccc  tcgacctgcc  cgccgccctc
        19751  gacgcccggg  ccgcccagcg  gctcgccgcc  gtgctgtccg  gcgccctcgg
        19801  cagcgacgac  cagctggcca  tccggccctc  cggggtcttc  acccgccgca
   10   19851  tcgtgcgggc  cgaggccacc  gccgggcggc  ccgccggcac  ctggacgccg
        19901  cgcggcacca  cactggtcac  cggcggctcc  ggcaccctcg  ccccgcacct
        19951  cgcccgctgg  ctggcccaac  gcggcgccga  gcacctggtc  ctgatcagcc
        20001  ggcgcggcac  ggccgccccg  ggcgccgccg  aactcgtcgc  ggaactggcc
        20051  gagtcgggca  ccgaggcgac  cgtcgccgcc  tgcgacatca  ccgaccgcga
   15   20101  cgcggtcgcc  gcgctgctgg  ccgacctcaa  ggccgacggg  cgcaccgtcc
        20151  gcaccgtcgt  gcacaccgcc  gccaccatcg  agctgcacac  cctggacgcc
        20201  accaccctgg  cggacttcga  ccgggtgctg  cacgccaagg  tcaccggcgc
        20251  ccaggtcctc  gccgaactgc  tcgacgacga  agagctggac  gacttcgtcc
        20301  tgtactcctc  caccgccggc  atgtggggca  gcggcgccca  cgccgcctac
   20   20351  gtcgccggca  acgcctacct  cgccgcgctc  gccgagcacc  gccgggccaa
        20401  cggactgccc  gccctgtcgc  tgtcctgggg  catctgggcc  gacgacctca
        20451  aactgggccg  ggtcgatccc  cagatgatcc  ggcgcagcgg  cctggagttc
        20501  atggacccgc  agctggccct  gagcggcctg  cagcgggcgc  tggacgacaa
        20551  cgagaacgtg  ctcgcggtcg  ccgacgtgga  ctgggagacc  taccaccccg
   25   20601  tctacacctc  cggccgaccc  accccgctct  tcgacgaggt  gccggaggtc
        20651  cgccggctca  ccgcggccgc  cgagcagagc  gccgggaccg  tcgccgaggg
        20701  cgagttcgcc  gccgcgctgc  gcgccctgtc  cgacgccgag  cagcagcgca
        20751  ccctgctgga  gaccgtccgc  accgaggcgg  cgtccgtcct  cgggctgtcc
        20801  tccgccgagg  acctcaccga  ccagcgggcc  ttccgcgacg  tcggcttcga
   30   20851  ctcgctgacc  gccgtcggcc  tgcgcaaccg  gctcgcctcc  gtcaccggcc
        20901  tgacgctgcc  ctcgacgatg  gtcttcgact  accccaaccc  ggccgcgctc
        20951  gccgcctatc  tgcacggcga  gctggccggc  gcccggtccg  ccgccgccgg
        21001  cgccgccgcc  gtccgaccg   gcgccccga   cgccgacgac  ccgatcgcga
        21051  tcgtcggcat  gagctgccgc  tacccggcg   gggtcggctc  cgccgaggac
   35   21101  ctgtggcgga  tcgccctgga  cgaggtcgac  gcgatctccg  gcttccccgc
        21151  cgaccgcggc  tgggacgccg  agggcctcta  cgacccggac  cccgaccggc
        21201  ccggccgcac  ctactccgtc  cagggcggat  tcctgcgcga  cgtcgccgag
```

```
      21251   ttcgacccgg  gcttcttcgg  gatctcgccg  cgcgaggcgc  tgtcgatgga

      21301   cccgcagcag  cggctcctgc  tggagaccgc  ctgggaggcg  ttcgagcacg

      21351   ccggcatcga  cccggtcggc  cagcgcggca  gccgcaccgg  caccttcgtc

      21401   ggcgccagct  accaggacta  cgcctccggc  gtgcccaaca  gcgagggctc

   5  21451   cgaaggccac  atgatcaccg  gcacgctctc  cagtgtgctg  tccggccggg

      21501   tgtcctacct  cttcggcttc  gagggccccg  ccgtcacgct  cgacaccgcc

      21551   tgctcctcct  ccctggtcgc  catgcacctg  gcctgccagt  ccctgcgcaa

      21601   cggggagagc  tcgctggccc  tggccggcgg  cgtcagcatc  atgtccaccc

      21651   cgatgtcgtt  cgtcggcttc  agccggcagc  gcgccctcgc  cgaggacggc

  10  21701   cgctgcaagg  cgtacgcgga  cggcgccgac  gggatgaccc  tcgccgaggg

      21751   cgtcggcctg  gtgctgctgg  agcggctgtc  cgacgcccgc  gccaacgggc

      21801   accaggtgct  cgccgtgatc  cgcggctccg  ccgtcaacca  ggacggcgcc

      21851   tccaacggcc  tgaccgcacc  caacggcccg  tcccagcagc  gcgtcatccg

      21901   ccaggcgctg  gccaactccg  cggtggcgcc  cggcgacatc  gacgtcctgg

  15  21951   agggccacgg  caccggtacc  gccctcggcg  accccatcga  ggcgcaggcc

      22001   ctgctcgcca  cctacggcca  ggaccgcgcc  cccgaacggc  cgctgctgct

      22051   cggctcggtg  aagtccaaca  tcggccacac  ccagatggca  tccggcgtcg

      22101   ccagcgtcat  caagctcgtc  cgcgccctcc  aggaaggcgt  ggtgcccaag

      22151   tccctgcaca  tcgaccggcc  ctccacccac  gtcgactggt  cctcgggcgc

  20  22201   catcgggctg  ctcaccgaac  gcaccccgtg  gcccgagacc  ggccggccgc

      22251   gccgcgccgc  cgtctcctcc  ttcggcatca  gcggcaccaa  cgtccacacc

      22301   atcctcgaac  aggcccccgc  ggacgaggcg  cccacgcccg  ccgacccgcc

      22351   gcgggacggc  ctggtgccgg  tcctgctctc  cggccgcggc  gaggccgcgc

      22401   tgcgcgccca  ggccgcccgc  ctgctcgcct  tcgtcgagga  gcggcccgag

  25  22451   gcccacctca  ccgacctcgc  ccactccctc  gccacctcgc  gcgccgcgct

      22501   ggaacgccgc  gccgccgtca  tcgccgccga  ccgcgacacc  ctgacccgcg

      22551   gcctgcgcgc  cctgtccgac  ggccggcccg  accccggcct  ggtccagggc

      22601   accgcgggac  gcggccggac  cgccttcctg  ttcaccggac  agggcagcca

      22651   gcgccccggc  atgggccgcg  aactccacga  ccgctacccg  gtgttcgccg

  30  22701   acgcgctgga  cgaggtgctg  gcccggctcg  acgacggacc  ggaccggccg

      22751   ctgcgcgagg  tgctgttcgc  cgcgcccgac  tccgccgagg  ccgcgctcct

      22801   ggaccggacc  ggctacgccc  agcccgcgct  gttcgccgtc  gaggtcgcgc

      22851   tgttccgcct  gctgacgtcc  tggggcctga  ccccggacta  cctggccggc

      22901   cactccgtcg  gcgaactcgc  cgccgcgcac  gtcgccggcg  tgctgtcgct

  35  22951   ggacgacgcc  tgcactctgg  tcgccgcccg  cggccggctc  atgcaggcgc

      23001   tgcccgaggg  cggcgcgatg  gtcgccctgg  aggccgcgga  ggacgaggtc

      23051   ctgccgctcc  tggagggggct  caccgaccgg  gtgtccgtcg  ccgccgtcaa
```

```
         23101  cgggccgcgg  tccgtggtcg  tcgccggcgt  cgaggaggac  gtgctcctcc
         23151  tcgccgacct  cttcgccgcc  gacgggcgcc  ggaccaagcg  gctgcgggtg
         23201  agccatgcct  tccactcgcc  gctgatggac  gccatgctcg  acgacttcgc
         23251  cgccgtcgcc  cgcgggctga  cctaccaccc  gccgacgatc  ccgttcgtgt
     5   23301  cgaacgtcag  cggcggcctg  gccaccgccg  aacaggtccg  cacgcccgac
         23351  tactgggtcg  ggcacgtccg  cgccgcggtc  cgcttcgccg  acggcatcga
         23401  ctggctcgcc  acccagggcg  acgtccacac  cttcctggag  ctcggcccgg
         23451  acggcgtgct  cagcgccatg  gcccgggaga  gcctcaccga  cccgtcccgc
         23501  acggcactgc  tgccgaccct  gcgcggcgac  cggcccgagg  aacctgccct
    10   23551  ggtcaccgcg  gtcgccgcgg  cccacgcgca  cggcgcccgc  gtcgactgga
         23601  gcgggtactt  cgccgaccac  ggcgcgcgcc  ggaccacgct  gccgacctac
         23651  gcgttccaac  gcgagcggta  ctggcccgac  accaccgccg  ccacgagcgc
         23701  ccacacgccc  ggatccgccc  tcgacgccga  gttctgggcc  gccgtcgagc
         23751  gggacgacgt  cgccgccctc  gccgcctccc  tggacctgga  cgacgccacc
    15   23801  gtcaccgcga  tggtccccgc  gctcaccgcc  tggcgccggc  gccgcggcga
         23851  gcagaccgaa  ctggactcct  ggcgctaccg  cgtcacctgg  aagccgcgcg
         23901  gcggcgccac  cgcacccgcc  gccctcaccg  gccgctggct  cgtgctcgtc
         23951  ccgcacgacc  accaggaccg  tcaggacgac  gcgaccgcgg  cctgggcagc
         24001  cgacgtcgag  accgccctgg  gcaccaccac  cgtccggctg  acggtcacca
    20   24051  ccaccgaccg  cgccgcgctg  gccgcccgga  tcaccgaagc  cgccggcgac
         24101  cagggcccgt  tcagcggtgt  gctgtccctg  ctgccgctcg  ccaccggcga
         24151  cgccggccac  cccggtgcgc  ccgccgccct  caccctcacc  accaccgccg
         24201  tccaggccct  cggcgacgcc  ggcatcgacg  cgccgctgtg  gaacgtcacc
         24251  cgcggggccg  tggccgtcgg  ccgcgccgaa  caggtcaccg  cgcccgaaca
    25   24301  ggccgccgtc  tggggcctgg  gccgcgccgt  cgccctggaa  ctgccggccc
         24351  ggttcggcgg  caccctcgac  ctgcccgcca  ccctggacgg  ccaggccgcc
         24401  cgccggttgc  gcgcggtgct  cgcggctacc  gacggcgagg  acgcggtggc
         24451  cctgcggccc  tccggcgtct  tcctccgccg  cctggcccac  gccccggccg
         24501  gccccgacac  cgcccgcacc  gccttcgacc  cggcggccgg  caccgtcctg
    30   24551  atcaccggcg  gcaccggcgg  catcggcggc  cacgtcgccc  gccgcctggc
         24601  ccgcgacggc  gccacccacc  tgctgctcac  cagccgccgc  ggcccggccg
         24651  cccccggcgc  cgacgcgctc  cgcgccgaac  tggaggaact  gggcgcccgg
         24701  gtcaccctcg  ccgcctgcga  cgccgccgac  cgcgacgcgc  tggccgcgct
         24751  cctcgccgaa  ctgcccgacg  acgccccgct  gtgcgcggtg  ttccacaccg
    35   24801  ccggcgtcgt  cgaggaccac  gtcgtggacg  cgctcacacc  ggagaacttc
         24851  gccgcggtgc  tgcgcgccaa  gaccgtcgcc  gcccaccacc  tgcacgagct
         24901  gaccgccgac  ctggacctcg  ccgctttcgt  gctgttctcc  tccacggccg
```

```
        24951  gcgtcctcgg cgccgccgga cagggcaact acgccgcggc caacgcccac
        25001  ctggacgccc tcgccgaaca ccgccgctcc cacggcctga ccgcgctgtc
        25051  cgtcgcctgg ggcccgtggg ccggctccgg catggtcgcc gacgccgccg
        25101  aactcaccga ccgggtacgg cgcggcggct tcgaaccgct cgcccccgaa
  5     25151  ccggccgtgc gcgccctgct gcgcgccatc gagaacgacg acaccaccgt
        25201  cgcgctcgcc gacatcgact gggagcgctt ccagcgcgcc ttcgccgcgg
        25251  tccgcccgct gccgttcgtc gccgacctcc ccgagaccgg ccgggccacc
        25301  cccgcgaccg ccaccggcgc cgccaccggc ctgcggcagc aactcgccga
        25351  actgccggag cacgagcgcc cggcagcggt cctggacctg ctgcgtaccc
  10    25401  aggtcgccgc cgtcctcggc cacgccgacc cgcgcaccgt cgaggacgac
        25451  cacgccttcc gcgacctggg cttcgactcg ctgaccatcc tggaactgcg
        25501  caacgccctc aacgccgcca ccggcctgag cctgcccgcc accctggtct
        25551  acgacctgcc caccccgcgc gagatggcgg acttcctgct cgccgaactc
        25601  ctcggcaccc tgcccaccga caccgccgcg accgtcgcca gcacggcctc
  15    25651  ccccaagctc tcagcttcgt tcgagcaggg cggtaccccc ttcgacgacc
        25701  cgatcgccgt catcggcatc ggctgccgct tccccggcgg cgtcaccacc
        25751  ccggaggagc tctggcagct cctcgacgag ggccgcgacg gcatcagccg
        25801  cttcccccgac gaccgcggct gggacctcgc cgcgctgggc gccggcgcct
        25851  ccgacaccct ggagggcggc ttcctgaccg gcgtcgccga cttcgacgcc
  20    25901  cggttcttcg gcatctcgcc ccgcgaggcg ctggccatgg acccccagca
        25951  gcggctgctg ctggagacca cctgggaggc gctggagcgg gccggcatcg
        26001  acccgaccac gctgcgcggc tccaccaccg gcgtcttcgt cggcaccaac
        26051  ggccaggact acccgacgct gttgcgccgc tccgcctcgg acgtggccgg
        26101  ctacgtcgcc accggcaaca ccgccagcgt gatgtccggc cgcctgtcct
  25    26151  acgcgctcgg cctcgaaggc ccggccgtca ccatcgacac cgcctgctcc
        26201  tcctcgctcg tcgccctgca ctgggccggc cgggcgctgc gcgccggcga
        26251  gtgcgacctc gtggtggccg gcggcgtctc ggtcatggcc agcccggact
        26301  ccttcgtcga gttctccacc cagggcggcc tggcacccga cgggcgctgc
        26351  aaggcgttct ccgacgccgc cgacggcacc gcctggtccg aaggcgtcgg
  30    26401  catcctcgtc ctggaacgcc tctccgccgc ccgccgcaac ggccaccagg
        26451  tcctcggcct gatccgcggc accgccgtca accaggacgg cgcgtccaac
        26501  ggcctgaccg cgcccaacgg cctctcccag cagcgcgtca tcgcccaggc
        26551  actcgccgac gcccgcctgc gccccgccga catcgacgcg atcgaggcgc
        26601  acggcaccgg caccaccctc ggcgacccga tcgaggcccg cgccctgatc
  35    26651  accgcctacg gccgggaccg ggacgccgaa cggccgctgc tgctgggcac
        26701  cgtcaagtcc aacatcgggc acacccaggc cgccgccggt gccgccggcg
        26751  tcatcaagat gctgatggcg atgcgccacg gcacccctgcc caggacgctg
```

```
26801   cacgtgggca   ccccgtccag   ccacgtcgac   tggagcggcg   gcaccgtcgc
26851   gctcctcgac   gacgcgcggc   cctggccacg   gaccgggcag   ccgcggcgcg
26901   ccggcgtctc   cgccttcggc   gtcagcggca   ccaacgccca   cgtcgtcgtc
26951   gagcaggccc   cggaaaccga   agcccccgcc   gccccggccg   ccgagccggc
27001   gccggaggcc   acgcccaccg   tcgtcccctg   ggtcgtctcc   ggacgcagcc
27051   gggaagcgct   ccaggcgcag   ctggaccggc   tcaccgcgca   caccgccgcc
27101   caccccgcgc   gctcggcggc   ggacgtcggc   cgctcgctgg   ccaccgaccg
27151   cacgctcttc   ccgcaccgcg   ccgtgctgct   cgccggcccg   gacggggtgc
27201   gcgaggccgc   ccgcgccgcc   gcgccccgca   cccccggccg   caccgcgttc
27251   ctgttctccg   gacagggcgc   ccagcacgcc   ctgatgggcc   acgacctgta
27301   ccagcgcttc   ccggtctacg   ccgacgcact   ggacaccgtc   ctcgcccagt
27351   tcgacaccgt   gctggacgtc   ccgctgcgcg   ccgcgctgtt   cgccgcgccg
27401   ggcacccccg   aggccgcgct   cctggaccag   accggcttca   cccagcccgc
27451   gctgttcgcc   gtcgaggtcg   cactgttccg   gctcgccgag   tcctggcggc
27501   tgacgccgga   cttcgtcgcc   ggccactcca   tcggcgagat   c
```

SEQ ID NO: 3

NysD2 (incomplete - probably C-terminally truncated).

```
  1   msftypvsmp   wlqgreldyv   teavgggwis   sqgpyvrrfe   eafaayndvp
 51   fgvacssgtt   altlalralg   vgpgdevivp   eftmiasawa   vtytgatpvf
101   vdcgddlnid   vsrieekitp   rtkvimpvhi   ygrqcdmdav   lnlayeynlr
151   vvedsaeahg   vrprgdiacf   slfankiisa   geggvclthd   phlaeqmahl
201   ramaftkdhs   flhkklaynf   rmtnmqaava   laqteqldti   lalrrdiekr
251   ydealrdipg   itlmpprdvl   wmydlraerr   delcaylage   gietrvffkp
301   msrqpgyfsa   dwpalnaarl   sadgfylpth   tgltaqeqef   itgri
```

SEQ ID NO: 4

NysD1

```
  1   vtlpsgntrl   gwrrrrmhsp   gdragrvrga   rarrpatfrg   vlsmganrrp
 51   ilfvsyaesg   llnpllvlag   elsrrdvadl.wfatdekard   evaavvdgsp
101   vrfaslgdtv   sqmsavtwdd   atyaevtqrs   rfkahaavir   hsfapesrma
151   kyrrleeive   evepalmvie   smcqfgyela   itkgipfvlg   vpfvpsnvlt
201   shvpfaksyt   psgfpvphsg   lpaamslaqr   ienqlfrlrt   lgmfltsdvr
```

```
      251  kvveednrvr  telgiapqar  qmmaridhae  qvlcysvrel  dypfpmhpkl
      301  rlvgtmvppl  pqapdddgls  dwlsaqksvv  ymgfgtitrl  treqvaslve
      351  varrldgrgh  qvlwklprgq  qellppaael  pdnlriegwv  psqldvlahp
      401  nvkaffthag  gngyheglyf  gkplvvrplw  vdcddqairg  qdfgvsltld
      451  rpetvdtedv  ldkitrvldq  psfteraehf  agllrdaggr  aaaadlllgl
      501  palatd
```

SEQ ID NO: 5

NysA


```
        1  mtigadedpv  vvvgmacryp  ggvagpedlw  elvrtgrdat  tafpddrgwd
       51  laalagdgpg  rsatreggfl  tgaadfdaaf  fgmspreavs  tdpqqrlvle
      101  tawealerag  idphslrgsr  tgvfvgasgq  dyaavthasp  ddldghaltg
      151  lapgvasgrl  ayvlglegpa  vtvdttssss  lvalhwavra  lragecstal
      201  aggvtvmstp  aafvghtrqg  glapdgrckp  fsddadgtaw  aegvgivvle
      251  hlstaraagn  pvlavlrgsa  vnqdgasdgl  tapsgpaqer  viraaladar
      301  lapadidlve  ahgtgtrlgd  pvearallaa  ygqdrdpdrp  lrlgslkstl
      351  ghaqaaagig  gviktvltlr  hglmprirhl  atptrqvdws  qgavapltdh
      401  tpwppadrpr  ragvssfgis  gtnahvilee  appadvpvtr  pgtlrpstvp
      451  wpvsaatpea  ldaqlarlra  hlrthsdldp  ldvgyslatg  raalrhravl
      501  lppadgtaad  avehargaah  qrrtavlfsg  qgsqrpgmgr  elaarfpvfa
      551  dalddalral  drhldgpvre  vmwgtdaall  drtgwtqpal  favevalhrl
      601  vaslgvtpdf  vgghsvgeia  aahvagvlsl  edacrlvaar  atlmqalpag
      651  gamaaleate  devapllgah  lalaavngpt  avvvagaeda  vrqltarfad
      701  rgrrtsrlav  shafhsplme  pmldafrdvv  srltfhqpsi  plvsnltgel
      751  agseitsaey  wvrhvrdtvr  fadgitalak  agadvlielg  pggvlsamar
      801  dtlgpdsttd  vvpalskgrp  eetafagalg  rlhtlgvpvd  wpafyagtga
      851  rrvelptyaf  qhvrhwptpp  rpngagpgal  ghpllgsave  ladgggtvcs
      901  galslrthpw  ladhtvagrv  vlpatallel  avragdeagc  dvlhelhltt
      951  ppalpddaal  hvqvhvgpad  ttgrravtvh  trpdhhpagd  wtrcatgtlg
     1001  stppsaaeaa  tggtpaawpp  adaepldlad  hyerladrgf  dygptfrglr
     1051  aawrrgaeif  advecppgta  ddapdhglhp  alldaarhaa  mavdgtvpva
     1101  whgvrlhavg  atalrvrirp  tttgtltlta  vdvhgapvvt  vealtarplt
     1151  deeraaprtp  rqargetpad  arparpaaar  pgpageplpd  ttgshptagh
     1201  laalppaare  rqlldlvrtq  aaavlghpgp  eavgtrsvfk  elgfdslagv
     1251  eladrltart  glrlpatlvf  nfptperaah  rlgellaata  pldpgaygee
```

```
    1301   ltrfeaivtn lpqdgperra vadrldaivs alrqnspaev pssdedidtv
    1351   svdrlldiid eefett
```

SEO ID NO: 6

NysB

```
       1   mqepqqgqpd qqekivdylr rvtsdlrrar rrigeleskd nepiaivgmg
      51   crlpggvnsp eslwdlvrsg gdaisgfpvd rgwdletltg ngdgssathe
     101   ggflydaaef daaffgispr eatamdpqqr lllevaweal eragiaptal
     151   rgsrsgvfvg syhwgapsad aatelhghal tgtaasvlsg rlaytlgleg
     201   pavtvdtacs sslvalhlaa qslrvgessl aviggvtilt epsvfvefsa
     251   qgglapdgrc kafsdaadgt gwaegvgvlv aerlsdaqrn ghpvlavlrg
     301   savnqdgasn gltapngpsq erviqqalar tgltpadida veahgtgtrl
     351   gdpieaqall atygqghtpd qplwlgslks nightqaaag vagvikmvma
     401   lrhghlpptl hadapsshvd wsagsvrllt egqqwpetgr prraavssfg
     451   isgtnahall eqaphpadta dagddaapte pagapaalpw ivsghspqal
     501   rdqaaalaar vetdpalrpq dightlhtar allerravvv apdraellaa
     551   thelaagrsa navvegladv egrtvfvfpg qgsqwvgmga qlldesavfa
     601   eriaecaaal aeftdwslvd vlrgvvgaps lervdvvqpa sfavmvslaa
     651   lwgsrgvlpd avvghsqgei aaavvsgals lrdgarvval rsqaigrala
     701   grggmmsval svdvleprlv efegrvsvaa vngprsvvva gepealdalh
     751   arltaddira rriavdyash shqvedlhee llevlaelap rtsevpffst
     801   vtgdwldtar mdagywfrnl rgrvrfadav adllaaeyra fvevsshpvl
     851   tmavldliee agvtavatgt lrrdqggagr fllsaaevfv rgvdvdwaga
     901   fegtgaarvd lptyafqrer ywntrtaadr tpadapmdae fwaaveqadv
     951   saltaalgtd edsvaailpg ltswrrarsq rttldswryr vtwtplaqvp
    1001   ratltgtwll vttdgiddtd vagalesyga evrrlvldee ctdravlrer
    1051   lagaedvtgi vsvlaaaedd aarhpgltrg laltvslvqa lgdaeatapl
    1101   wfltrgafat gpsdpvtrpl qsqiagvgwt talehpqrwg gtvdlpdtld
    1151   araaqrlaaa lsgalgaedq lavraagvla rrivraghra grpartwapr
    1201   gttlitggsg tlapqlarwl aergaehvvl vsrrgadapg apeliaeaae
    1251   sgtevtvaac ditdrdavaa lladltadgr tlrtvihaaa aielsaladt
    1301   tvaefadvvh akvtgarild ellddaeldd fvlysstagm wgsgvhaayv
    1351   agnaylsala eqrrarglrt tsihwgkwpd drareladph rirrsgleyl
    1401   dpelaltalq hvldddetvi glmdidwdty hdvftagrpa hlfdqipevr
    1451   rrldqasvpd pagpaadgla arlhglaaae qdrllltlvr teaaavlgha
```

SUBSTITUTE SHEET (RULE 26)

```
      1501  saesfperra frdlgfdsvt avdlrnrlva gtglrlpstm vfdhpncaal
      1551  aaflkttalg vpgaapqqha atgtpadddp iavigmscry pggaatpeel
      1601  lrlaldgadv isefpadrgw darglydpdp drpghtysvq ggflheaagf
      1651  dpgffgispr eavamdpqqr llletsweaf eragidpasl rgsaagtffg
  5   1701  asyqdysstv qngtgeseah mvtgtaasvl sgrvsyllgl egpavtvdta
      1751  cssslvalhl acqslrdges slalaggaav matphafvgf srqralakdg
      1801  rckpfsdtad gmtlaegvgv vllerlshar anghrvlavi rgsavnqdga
      1851  sngltapngp sqqrvirqal anagltgadv daveahgtgt klgdpieaqa
      1901  llatygqdrd aerplllgsv ksnightqaa agvagvikmv lamdagelpg
  10  1951  tlhldapssh vdwtagavel lrgrtpwpes grprragvss fgisgtnahl
      2001  ileqapatep padpdrlrdt atdtvvpwpl aakspaalra qaarllatve
      2051  hdpdlppapv ghalattraa lehravvvge rredflrgla alstgastag
      2101  lvsgiagpdp egavfvfpgq gsqwwgmgre llatsevfrt aiddcatala
      2151  pyvdwslhdv lagegdpall ervdvvqpal fammvglsal wrshgvvpaa
  15  2201  vvghsqgeia aacvagalsl adaarvvalr sqalpqlsgr ggmmsvsapv
      2251  ervtallapw qealsvaavn gpssvvvsgd tdaldalhta cqeqgvrark
      2301  vsvdyashgr hveavrdela rvlapvdpra pevpfystvt gdrvddaafd
      2351  gaywytnlrq tvrmeeatra llaaghrvfi evsphpvlaa piqetqeava
      2401  eatggsavvl gslrrdeggp rrfltslaea hthgapvdwt ttfarsayqp
  20  2451  vdlptypfqr qdfwpearpa tpaagadasd aafwqlvenq dlaaladalg
      2501  vpaddehtal gtvlpalsaw rakaqartri delryhvqwt rvaepaaapt
      2551  tgrllvavpp dhadapwvaa aldalgtdtv rfeakgtdra gwaaqiaqlv
      2601  edgeeftgvv sllaaaedlh pdfgsvplgl gqtlvlvqal gdagltaplw
      2651  cltrgavatg rddaldsptq galwglgrvv alehpdrwgg lidlpatlda
  25  2701  raaarltgll adpagedqla vratgvlarr mvhaapsapr tgrrwrgrgt
      2751  clitggtggi ggrvarwmae hgaahlvlts rrgpdapgaa alraelealg
      2801  arvtlaacdv adrdalaall adlpadqplt svfhsagvad gdaraadltl
      2851  dqldallrak ltaahhlhel tapldldafv lfssgaavwg sggqpgyaaa
      2901  nayldalaah rrsldlpgas vawgtwgevg matvpevher lhrqgvrame
  30  2951  pdhaigalqq mledddttla vtlmdweafa psftatrpsa lfstvpeavr
      3001  avtgdpgtta gddvdsatpp lrrhleelsa aergralvea vraeasatlg
      3051  hdtpdaipag rafrdvgfds vtavelrnrl rtalglplpa alvfdhptpt
      3101  alaghlgall fgtapedagt grpddpdari realatvpig rlrkaglldm
      3151  vlkladgdat dapapeadap seslddmdae allrlatens an
```

SEQ ID NO: 7

NysC

```
  5       1  mstnpdkyve  alrsslkeie  rlrrqneqlv  aaavepvavv  gigcrfpggv
         51  tspedlwelv  aegrdvigpf  pqdrgwdlek  lagggeggsl  aqvggfveda
        101  agfdpgffgi  spreavamdp  qqrilleitw  ealeragidp  stlrgtptgv
        151  fvgttgqdyg  evikasaedv  evysttghaa  svisgrlsyt  lgaegpavtv
        201  dtgcssslva  lhwavqalrg  gecsmalagg  asimatpgpf  vaftaqsgla
 10     251  adgrckpfsd  radgtgwgeg  agmlvlmrls  daqregrpvl  avlrgsainq
        301  dgasngltap  ngpsqqrvir  aaldsahlta  adidaveahg  tgttlgdpie
        351  aqallatygq  drprplwlgs  vksnightqa  asgaagvikm  imalqrgvlp
        401  rslhateptt  dvdwtagsvd  lldetvawpe  tgrarragvs  sfgisgtnah
        451  vileqaptap  eeptteptvr  pavvpwalsa  rtaaaldaqr  arltghladt
 15     501  pdadpldvgy  aladgratfe  hravllpdgt  elahgtageg  pcavlfsgqg
        551  sqrpgmgrel  harfpvfaaa  fdeitalldt  hldrplrevv  wgtdadllnd
        601  tgwaqpalfa  vevalyrlva  slgvtpdfvg  ghsigelaaa  hvagvlsled
        651  actlvaarar  lmqalprgga  mlairatede  vtphltddvs  iaavngptsv
        701  vvagteeava  aigarftaqd  rkttrlrvsh  afhsplmdpm  laefravaag
 20     751  ltyhepripv  lsnltgtvaa  vadlcsadyw  vrhvreavrf  adgvtaltdr
        801  gvttlvelgp  dgvlsamaqe  slpdgaaavp  llrkdrpeel  savtglarah
        851  vrgvtvrwag  lfdgtgarra  dlptypfqhq  rfwptaaraa  qdvtaaglga
        901  adhpllgatv  eladgagylf  tsrlsvrthp  wladhgvqgr  allpgtafve
        951  lavragdeag  cdrveeltla  aplvlpergg  vqlqvrvgap  daagrrtlgi
 25    1001  fsrvedgfdl  pwsqhatgvl  tagagapdpt  fdatvwppsg  aepvdltgay
       1051  erlaalgfqy  gpafqglraa  wrrdtevyae  valpdgadtd  paafglhpal
       1101  ldaaqhaaay  adlgaisrgg  lpfawegvsl  aaagattvra  riapagedtv
       1151  tiavydaagg  tvlsvdslvs  revpadapga  agtvhrdslf  hvewtplqgr
       1201  pgpapatvav  lgpdpdalad  tlratgirtt  aprdlaalad  aegpvpdlvv
 30    1251  ttltttpgap  vpdaahatta  avlalaqqwl  addrfadarl  vlvtrgatdg
       1301  tdpaaaaagg  lirtartenp  grfallldlap dtgrpdpetl  atalaashde
       1351  pdlavrgtdv  haarlarvpl  atepttwnpd  gtvlitggtg  glgavlarhl
       1401  vathgvrhll  lasrrgpaad  gaddltaelt  glgatvhiaa  cdvadpaala
       1451  dllgtvpagh  pltvvvhtag  vvddgvlgsl  tpqrldtvlr  pkadaawhlh
 35    1501  eatrhldlda  fvlfssvaat  lgspgqanya  agnafldala  arraatglpa
       1551  tslawgpwtq  svgmtsslsd  ldveriarsg  mppltleqgt  alfdaalaag
       1601  paalapvrld  lpvlrtqgdi  aplirglirt  pvrrtaaqvs  qtadglaqrl
```

```
         1651  agldaaarre  allelvrtqi  aqvlghadat  evetgrqfqd  lgfdsltave
         1701  lrnalntatg  lrlpatmvfd  yptphaladh  lrdellgtea  esttavpvpt
         1751  rtagtddpiv  ivgmacrypg  giaspedlwr  lvsqgadatg  pfptnrgwdl
         1801  dnlydpdpdr  pgrthvragg  flhdagsfda  dffgmsprea  matdsqqrll
    5    1851  lelsweaver  agidpaslrd  sgtgvfagvm  yndygttltg  deyeafrgng
         1901  sapsvasgrv  sytlglegpa  vtvdtacsss  lvalhwaaqa  lragecslal
         1951  aggvtvmstp  stfvefsrqr  glapdgrska  faeaadgvaw  segvgmlvle
         2001  rqsdavrngh  eilavvrgsa  vnqdgasngl  tapngpsqqr  virqalasgg
         2051  lstadvdave  ahgtgttlgd  pieaqallat  ygrdrdpenp  lllgsiksni
    10   2101  ghtqaaagva  gvikmvmamr  hgvlpqtlhv  dapsshvdws  vgavellteq
         2151  tawpetgrar  ragvssfgis  gtnahvvieq  sptavpatpa  sadrsveepp
         2201  avpwalsgkt  pdalrdqaar  llahveahpa  lrpvdisysl  iatrtafdhr
         2251  avvlgtdrae  alraltalaa  getdpaaltg  tvrtgrtafl  fsgqgsqrlg
         2301  mgrvlyerfp  afaealdtvl  taldaelghp  lrdiiwgeda  qlvdrtgytq
    15   2351  palfaieval  frlleawgit  pdfvaghsig  eiaaahvagv  lslgdacrlv
         2401  varavlmqsl  peggamiavq  atedevlpll  tddvsiaavn  sptsvvvsgy
         2451  enatlavarh  fadqgrrttr  lrvshafhsp  lmapmlddfr  avvesltfta
         2501  pttpvvsnlt  gelapaealc  sadywvrhvr  eavrfadgir  tladrgvttf
         2551  velgpdsvls  amaqesapeg  agtipllrrd  rpeeqavlaa  lchlqvlgve
    20   2601  adwsatfrgl  dpvrvdlpty  afqhrwfwpa  arparpddvr  aaglgaaehp
         2651  llgaavqlpd  ddgalftgrl  slrthpwlad  htvlgtvllp  gtalvelavr
         2701  agdetgsghl  eeltlaaplt  lpedgatllq  vrvgsaddtg  rrtvtvharp
         2751  ddtadrtwtl  hatgvlattp  paaaafdttv  wppadaeplt  tddcyahftt
         2801  hrfaygpafq  glraawragd  vlyaevalpe  satdeaaafg  lhpalldagl
    25   2851  haalladdrd  tglpfswegv  tlhasgatal  rvrlapngpn  glsvtaadpa
         2901  gnpvatvtrl  larpldaeql  tihsaltrda  lfhldwtpvp  lpdtansapp
         2951  allgpdtavl  adalgdpava  rhatlddlla  gdttppatvl  vplgapldgd
         3001  taqhahaltr  saltlvqqwl  atdrladsrl  vfvthgavat  ddapptdlaa
         3051  aavwglirsa  qtenpgtftl  ldldtepdst  talsraltld  epqlllragr
    30   3101  araarltrtp  apttthtpw  sadgtvlvtg  gtgglgglva  rhlvrscgvr
         3151  hllltsrsgv  gaagaaglva  eleslgarvv  vaacdvgdgs  avaelvagvs
         3201  esyplsavvh  aagvlddgvv  gsltperlaa  vlrpkvdgaw  nlheatrgld
         3251  ldafvvfssv  agvfggagqa  nyaagnafld  almvhrvagg  lpgvslawga
         3301  wdqgvgmtag  lterdvrraa  esgmplltvd  qgvalfdaal  atgsaalvpv
    35   3351  rldlaalrtr  gdiapllrgl  vraplrrtaa  tglatgadtg  lvqrlgrldh
         3401  aqrhealldm  vrssaalvlg  hadgnaidae  rafrdlgfds  ltavelrnrl
         3451  rtatglhlsa  tmvfdhptls  alaehlrdel  fgavesevrv  pvqalpptad
```

SUBSTITUTE SHEET (RULE 26)

```
      3501  dpivvvgmac  rfpggvtspe  dlwrlvddgt  daittfptnr  gwdldnlydp
      3551  dpehfgtsyt  rsggflheag  efdpaffgms  prealatdsq  qrlllesswe
      3601  aieragidpl  tlrgsatgvf  agvmysdygs  ilggkefegf  qgqgsagsva
      3651  sgrvsyalgf  egpavtvdta  cssslvalhw  aaqalragec  slalaggvtv
   5  3701  mstpstfvef  srqrglapdg  rskafaeaad  gvgwsegvgi  lvlerqsdav
      3751  rngheilavi  rgsavnqdga  sngltapngp  sqqrvirqal  asgglstadv
      3801  daveahgtgt  tlgdpieaqa  llatygrdrd  penplwlgsl  ksnightqaa
      3851  agvagvikmv  mamrhgvlpq  tlhvdapssh  vdwsvgavel  lteqtawpet
      3901  grvrragvss  fgisgtnahv  iveqpalves  paaepsgrep  gvvplplsgk
  10  3951  spealrdqaa  rllaglaerp  alrpldlgys  lattrsafdh  ravvlatdra
      4001  davraltala  aadadlsavv  gdtrtgrhav  lfsgqgsqrl  gmgrelyerf
      4051  pvfaealdva  idhldaalpa  qaslrevmwg  ddvelldetg  wtqpalfave
      4101  valfrlvesw  gvrpdfvagh  sigeiaaahv  vgvfsledac  rlvaaratlm
      4151  qalptggami  aiqaaedevt  qhltddvsia  avngptsvvv  sgaesaartv
  15  4201  adrlaengrk  ttrlrvshaf  hsplmdpmla  efravaegls  yatptlpvvs
      4251  nltgrlatad  dlcsaeywar  hvreavrfad  gvstlenegv  ttflelgpdg
      4301  vlsamaqqsl  tgdaatvpal  rkdrdeetsa  ltalahlhta  glrvdwaaff
      4351  agsgatrvdl  ptyafqhaty  wptgtlptah  aaavgltaae  hpllngsvel
      4401  aegegvlftg  rlslqshpwl  adhavmgqvl  lpgtallela  fragdeagcd
  20  4451  rveeltlaap  lvlpergavq  tqvrvgvadd  tgrrtvtvhs  rpehatdvsw
      4501  tqhatgtltm  gsapadtgfd  atawppadae  platddcyar  fttlgfaygp
      4551  vfqglraawr  agdvlyaeva  laestgdeat  afglhpalld  aalhaslvah
      4601  egeesngglp  fswegatlya  tgatalrvrl  tptgtdgrsv  aiavadtagr
      4651  pvaaidnlvs  rrvsgdqltg  aaglardalf  tldwnpvpen  lvpenpvpen
  25  4701  tggghaqdqd  grpaaatval  vgadgtaiaa  dltaagihtt  lhpdlttlat
      4751  tdadvpktvl  ipltgtgtgt  gtgtestdgi  gtgaaesdas  apspaevaht
      4801  lstaalalvq  ewtaqerfag  srlafvttga  taaggtdvmd  vaaaavwglv
      4851  rsaqseapdt  fvlidrdpgp  agthdrtaaa  ergqlllral  htdepqlalr
      4901  dggvlaarla  rfdtaaaltp  padrawrlds  takgslngla  ltpypaalap
  30  4951  ltghevrvev  raaglnfrdv  lnalgmypgd  dvgsfgseaa  gvvvevgpev
      5001  tglapgdqvm  gmitgsfgsl  avddarrlar  lpedwswetg  asvplvflta
      5051  yyalkelggl  ragekvlvha  gaggvgmaai  qiarhvgaev  fatasegkwd
      5101  vlrslgvadd  hiassrtldf  eaafaevagd  rgldvvlnsl  agdfvdasmr
      5151  llgdggrfle  mgktdiraad  svpdglsyqs  fdlawvvpet  igtmlaelmd
  35  5201  lfrtgalrpl  pvrtwdvrha  kdafrfmsma  khigkivltl  prswkpegtv
      5251  lvtggtgglg  glvarhlvrs  cgvrhlllts  rsgvgaagaa  glvaeleslg
      5301  arvvvaacdv  gdgsavaelv  agvsesypls  avvhaagvld  dgvvgsltpe
```

SUBSTITUTE SHEET (RULE 26)

```
        5351  rlaavlrpkv  dgawnlheat  rgldldafvv  fssvagvfgg  agqanyaagn
        5401  afldalmvhr  vagglpgvsl  awgawdqgvg  mtaglterdv  rraaesgmpl
        5451  ltvdqgvalf  daalatgsaa  lvpvrldlaa  lrtrgdiapl  lrglvkapir
        5501  raaattpgdt  glaeqltrlq  raerrdtlla  lvrdqaamvl  ghtsgdgvdp
    5   5551  srafrdlgfd  sltavelrnr  igaatglrlp  atavfdypta  dalaahllte
        5601  llgpdaesdp  depgdptagp  tddpiviigm  scrfpgdigs  pedlwrllgd
        5651  gadvvtdfpt  nrgwdldnly  dpdpahagts  yartggflhd  aadfdadffg
        5701  mspreamatd  sqqrlllless weaieragid pltlrdsrtg  vfagvmysgy
        5751  gtrldgaefe  gfqgqgsals  vasgrvsytf  gfegpamtvd  tacssslval
   10   5801  hlaaqalrgg  ectlalaggv  tvmsipdtfi  efsrqrglap  dgrskpfses
        5851  adgvgwsegv  gmlllerqsd  avrnghqila  vvrgsavnqd  gasngltapn
        5901  gpsqqrvirq  alasgglsta  dvdaveahgt  gttlgdpiea  qallatygrd
        5951  rdpenplllg  siksnlghtq  aaagvagvik  mvmamrhgvl  prslniteps
        6001  shvdwsagav  ellteqtawp  etgrarragi  ssfgisgtna  hvileqpeaa
   15   6051  rhsapeeadt  aeaaakapat  ahlpvmpwal  sgktpealra  qaarllahlq
        6101  qrpelapadi  alslatqrsq  fthravvlst  drdeatrals  alattaasdp
        6151  saltgtvtmg  rcavlfsgqg  sqrlgmgrel  yerfpvfaea  ldvvidhlda
        6201  alpaqaglre  vmwgddvell  netgwtqpal  faievalfrl  veswgvrpdf
        6251  vaghsigeia  aahvvgvfsl  edacrlvaar  atlmqalpag  gamiavqate
   20   6301  deviphltde  vaiaavngpt  svvisgaeea  tqtvaqhfad  qgrrttalrv
        6351  shafhsplmm  laefravaeg  lsyatptlpv  vsnltgqvat  adelcsaeyw
        6401  vrhvreavrf  adgvtaleae  gvrtflelgp  dgvlaamare  tvaddtvtvp
        6451  vlrrnmpeer  tlltalgrlh  ttgtpidwaa  llaptgarpv  dlptyafqhr
        6501  pfwpsgprdt  adaaavgiag  ashpllngiv  eladeegllf  tgrlslqshp
   25   6551  wladhavmgq  vllpgtalle  lalragdevg  cdhveeltla  aplvlperga
        6601  vqtqvrvgva  dttgrrtvti  hsrparattt  dsdthtgtdt  pwtqhatgvl
        6651  vaglpatatv  pfdatvwppa  haepvdladf  yasragegfg  ygpafqglra
        6701  awrrdgevfa  dvalpeagrt  eaeayglhpa  lldaglhaaw  lvapdgeptr
        6751  tgsvpfswrg  vflaasgass  vrvrlgrdsd  gtlslaiadt  tgapvasvqa
   30   6801  lsmrtvsvta  lsataglard  alfrldwasa  pepacqpddt  vtvipavavv
        6851  gtetseltse  ltaalraaga  dvdvrttlst  depapalial  plvasdqtgt
        6901  aeaapvpaav  hdltrralal  vqtrlqeqhf  adtkfvfvtr  gatvgrdvaa
        6951  aavwglvrsa  qsenpgcfal  vdldpdgavg  aaalvaalvs  gepqlavrgd
        7001  vlrvarlvrr  pltevgagad  gtgdgvgggs  gvsfsgegav  lvtggtgglg
   35   7051  avlarhlvae  ygvrdllllvs rsgeravgag  elvaelagvg  arvrvvacdv
        7101  tdraavvelv  gghavsavvh  aagvlddgmv  galtgerlsa  vlrpkvdavw
        7151  hlheatrgld  ldafvvfssl  agvfgspgqa  nyaaanafld  almtrrraeg
```

```
       7201  lpglslawgp  wsltdgtsgm  ladaeadrlt  rsgvppltae  qglalfdaal
       7251  atgdatcvpv  rldlsalraq  gevppllrsl  irgrsrraaa  aesatatglr
       7301  erlvglnpve  rqevllvdlvr  gqvalvlgha  daddvhpara  frelgfdslt
       7351  svelrnrlnt  vtglrlpatm  vfdyptvevl  vsyvldellg  tdaevatvqp
   5   7401  aavavaddpi  vivgmacryp  ggvaspddlw  rlvtdgvdav  spfptnrgwd
       7451  veslyhpdpd  hlgtsytrsg  gflheagefd  pgffgmspre  alatdsqqrl
       7501  llessweaie  ragidpvslr  gsrtgvfagv  mysdysamla  spefegfqgs
       7551  gsspslasgr  vaytlglegp  avtvdtacss  slvamhwamq  alrsgecgla
       7601  laggvtvmst  pavfvdfarq  rglspdgrck  afadaadgvg  wsegvgvlvl
   10  7651  erqsdavrng  heilavvrgs  avnqdgasng  ltapngpsqq  rvirqalasg
       7701  gltagdvdvv  eahgtgttlg  dpieaqalla  tygrdreper  plllgsvksn
       7751  lghtqaaagv  agvikmvlam  rhgvvprtlh  vdapsshvdw  segavellse
       7801  qaawpetgrv  rragvssfgi  sgtnahvile  rpeaarrpvm  etntvepstv
       7851  pwvlsgktpe  alraqaakll  ssieerpelr  lvdvgmslvt  grstfehrav
   15  7901  vlaadradaa  ralsaiaade  adaaaatgrv  gagrhavlfs  gqgaqrlgmg
       7951  relyerfpvf  aealdvvvdh  ldaalpaqag  lrevmwgdda  ellnetgwtq
       8001  palfaieval  frlveswgvr  pdfvaghsig  eiaaahvagv  fsledacrlv
       8051  aaratlmqal  paggamiavq  atedevtphl  tddvaiaain  gpnalvvsgv
       8101  edaaveigar  faaegrrttr  lhvshafhsp  lmdpmlaefr  vvaeglsyaa
   20  8151  pslpvvsnlt  gqvatadelc  saeywvrhvr  eavrfadgvt  aleaegvrtf
       8201  lelgpdgvla  amagasltes  slavpllrkd  rpeepaalaa  laqlhiagar
       8251  vdwpvlfagv  gagrvelpty  afqrgwfwpv  grvgvggdvg  avglgsaghp
       8301  llgaavelaa  gagvvltgrl  slsshgwlad  havmgrvfvp  gtallemvmr
       8351  agdevgcgrv  eeltlaaplv  lperggvrvq  vavdapdaag  rrgvgvyscp
   25  8401  dgvgqavwsq  havgvlasgv  adqvggfgdg  gvwppqgavs  vdaegcyelf
       8451  adagfgygpv  fqglravwrr  geelfaeval  sdevaesadt  atgfglhpal
       8501  ldaslhasll  sslegqsadg  gpalpfaweg  vslfasgata  lrvrlapage
       8551  havsvtavdp  tgapvisida  lrtrrltlde  vnashtqlsd  alfgvqwttv
       8601  pstpaadhps  vaiigtdhlg  laealsssssa  gatttttaaa  yesldaliaa
   30  8651  gpevsvpdvt  liglttedai  aqyvndhdat  vagqgtigag  aaavdaarrl
       8701  taealrtiqa  wladerlaar  rlvfvtrgaa  dgqdvaaaav  qglvrsaqte
       8751  npgtfglldl  dgteastavl  gealtsdepq  lllrdghlha  arltrlaspa
       8801  dtavptewna  dgtvlitggt  gglgaqfarh  lvdrygvrnl  llvsrrgpda
       8851  pgttelvael  tahgaevavq  acdvadgdav  aalvagvpde  hplravvhta
   35  8901  gvlddgvigs  lteerlatvl  rpkadaawhl  heatrgldld  afvvfssvag
       8951  vfggagqany  aaanafldal  maqrraaglp  glslawgpwd  qtggmtgmls
       9001  daeadrlars  gipplsaeqg  lalfdaalal  agtstpdraa  gsaaastsgt
```

```
      9051   gdtiaipaaa  lvapvrldla  alaaqgevpa  ilrglvrtrt  rrtaaggsvt
      9101   vaglvnrlsg  ltaderrqel  lelvrtqaal  vlghadpasv  dstaqfrdlg
      9151   fdsltavelr  nrlstatglr  ltatlvfdyp  ntdalaehlr  delfgavese
      9201   vrvpvqalpp  taddpivvvg  macrfpggvt  spedlwrlvd  agtdaittfp
 5    9251   tnrgwdlesl  ydpdpahlgt  sytrsggflh  eagefdpaff  gmsprealat
      9301   dsqqrllles  sweaieragi  dpltlrgsat  gvfagvmysd  ygsilggkef
      9351   egfqgqgsag  svasgrvsyt  lgfegpavtv  dtacssslva  lhlaaqalra
      9401   gectlalagg  vtvmstpgtf  vefsrqrgla  pdgrskafae  aadgvgwseg
      9451   vgilvlerqs  davrngheil  avirgsavnq  dgasngltap  ngpsqqrvir
10    9501   qalasgglst  advdaveahg  tgttlgdpie  aqallatygr  drdpenplll
      9551   gsiksnlght  qaaagvagvi  kmvmamrhgv  lpqtlhvdap  sshvdwsvga
      9601   vellteqtvw  petgrvrrag  vssfgisgtn  ahvileqpea  vqrlapgaae
      9651   tvepvaikps  aepslvpwal  sgkspealra  qaarlrdfla  erpeprsidi
      9701   ghslavtrsq  fdhraivlvd  dakapadsla  alaalasgva  dpavvsdavs
15    9751   tggsavlftg  qgaqrlgmgr  elygrfpvfa  ealdvvvdhl  daalpaqagl
      9801   revmwgddve  llnetgwtqp  alfavevalf  rlverwgvrp  dfvaghsige
      9851   iaaahvagvf  sledacrlva  aratlmqalp  tggamiavqa  tedevtphlt
      9901   devaiaavng  ptsvvisgae  eatqtvaqhf  adqgrrttal  rvshafhspl
      9951   mdpmlaefra  vaeglsyatp  slpvvsnltg  wlatadelcs  aeywvrhvre
20   10001   avrfadgitt  leaegvrtfl  elgpdgilsa  laqqslagea  vtvpvlrkdr
     10051   geestaltar  ahlhtrglie  dwqdffagvg  agrvelptya  fqrgwfwpvg
     10101   rvgvggdvga  vglgsaghpl  lgaavelaag  agvvltgrls  lsshgwladh
     10151   avmgrvfapg  tallemvmra  gdevgcgrve  eltlaaplvl  perggvrvqv
     10201   avdapdaagr  rgvgvyscpd  gvgqavwsqh  avgvlasgaa  dqvggfgdgg
25   10251   vwppqgavsv  daegcyelfa  dagfgygpvf  qglravwrrg  eelfaevals
     10301   devaesadta  tgfglhpall  daslhaslls  slegqsadgg  palpfawegv
     10351   slfasgatal  rvrlapageh  avsvtavdpt  gapvisidal  rtrrltldev
     10401   nashtqlsda  lfgvqwttvp  stpaadhpsv  aiigtdpfgl  adglsdalpl
     10451   veergdlaal  aasehpvpdl  vlvpvagtrr  tgvpadaegh  tdagtsdmlr
30   10501   svreataqvl  eqiqqwladd  rfeaarlvfv  trgavsvgeg  giadlaasav
     10551   wglvrsaqse  npgcfglldl  dldlaldsdl  apevdierdr  drdpvggtvq
     10601   palaaalhat  adepqlalrg  gtvqaarltr  ipapqtdrae  tdpaetdrpe
     10651   idtrrpgtvl  itggtgglgg  llarhlvaer  gvrslvlasr  sglaaegaek
     10701   lvadlealga  vvavqtcdva  dgdavaalva  gvsdeyplta  vvhtagvldd
35   10751   gvigslteer  latvlrpkad  aawhlheatr  dldldafvvf  sslagvlgga
     10801   gqanyaaant  fldalmaqrr  aaglpgvsla  wgpwdraggm  tgtlsdaead
     10851   rlarsgvppi  saeqglalyd  aatagerplv  vpvrldlaal  rglgdvpall
```

SUBSTITUTE SHEET (RULE 26)

```
        9051   gdtiaipaaa  lvapvrldla  alaaqgevpa  ilrglvrtrt  rrtaaggsvt
        9101   vaglvnrlsg  ltaderrqel  lelvrtqaal  vlghadpasv  dstaqfrdlg
        9151   fdsltavelr  nrlstatglr  ltatlvfdyp  ntdalaehlr  delfgavese
        9201   vrvpvqalpp  taddpivvvg  macrfpggvt  spedlwrlvd  agtdaittfp
    5   9251   tnrgwdlesl  ydpdpahlgt  sytrsggflh  eagefdpaff  gmsprealat
        9301   dsqqrllles  sweaieragi  dpltlrgsat  gvfagvmysd  ygsilggkef
        9351   egfqgqgsag  svasgrvsyt  lgfegpavtv  dtacssslva  lhlaaqalra
        9401   gectlalagg  vtvmstpgtf  vefsrqrgla  pdgrskafae  aadgvgwseg
        9451   vgilvlerqs  davrngheil  avirgsavnq  dgasngltap  ngpsqqrvir
   10   9501   qalasgglst  advdaveahg  tgttlgdpie  aqallatygr  drdpenplll
        9551   gsiksnlght  qaaagvagvi  kmvmamrhgv  lpqtlhvdap  sshvdwsvga
        9601   vellteqtvw  petgrvrrag  vssfgisgtn  ahvileqpea  vqrlapgaae
        9651   tvepvaikps  aepslvpwal  sgkspealra  qaarlrdfla  erpeprsidi
        9701   ghslavtrsq  fdhraivlvd  dakapadsla  alaalasgva  dpavvsdavs
   15   9751   tggsavlftg  qgaqrlgmgr  elygrfpvfa  ealdvvvdhl  daalpaqagl
        9801   revmwgddve  llnetgwtqp  alfavevalf  rlverwgvrp  dfvaghsige
        9851   iaaahvagvf  sledacrlva  aratlmqalp  tggamiavqa  tedevtphlt
        9901   devaiaavng  ptsvvisgae  eatqtvaqhf  adqgrrttal  rvshafhspl
        9951   mdpmlaefra  vaeglsyatp  slpvvsnltg  wlatadelcs  aeywvrhvre
   20  10001   avrfadgitt  leaegvrtfl  elgpdgilsa  laqqslagea  vtvpvlrkdr
       10051   geestaltar  ahlhtrglie  dwqdffagvg  agrvelptya  fqrgwfwpvg
       10101   rvgvggdvga  vglgsaghpl  lgaavelaag  agvvltgrls  lsshgwladh
       10151   avmgrvfapg  tallemvmra  gdevgcgrve  eltlaaplvl  perggvrvqv
       10201   avdapdaagr  rgvgvyscpd  gvgqavwsqh  avgvlasgaa  dqvggfgdgg
   25  10251   vwppqgavsv  daegcyelfa  dagfgygpvf  qglravwrrg  eelfaevals
       10301   devaesadta  tgfglhpall  daslhaslls  slegqsadgg  palpfawegv
       10351   slfasgatal  rvrlapageh  avsvtavdpt  gapvisidal  rtrrltldev
       10401   nashtqlsda  lfgvqwttvp  stpaadhpsv  aiigtdpfgl  adglsdalpl
       10451   veergdlaal  aasehpvpdl  vlvpvagtrr  tgvpadaegh  tdagtsdmlr
   30  10501   svreataqvl  eqiqqwladd  rfeaarlvfv  trgavsvgeg  giadlaasav
       10551   wglvrsaqse  npgcfglldl  dldlaldsdl  apevdierdr  drdpvggtvq
       10601   palaaalhat  adepqlalrg  gtvqaarltr  ipapqtdrae  tdpaetdrpe
       10651   idtrrpgtvl  itggtgglgg  llarhlvaer  gvrslvlasr  sglaaegaek
       10701   lvadlealga  vvavqtcdva  dgdavaalva  gvsdeyplta  vvhtagvldd
   35  10751   gvigslteer  latvlrpkad  aawhlheatr  dldldafvvf  sslagvlgga
       10801   gqanyaaant  fldalmaqrr  aaglpgvsla  wgpwdraggm  tgtlsdaead
       10851   rlarsgvppi  saeqgglalyd  aatagerplv  vpvrldlaal  rglgdvpall
```

```
10901   rglvrtparr  taaagaapsa  dvltrqlagl  ggaeqeevll  rlvrgqaavv
10951   lghadgsaig  agrqfqelgf  dsltavefrn  rlnaatglrl  patllfdypt
11001   padvvghlrg  rlgtgevsga  gsvlaaldnl  eaviaglsld  dagehqlvag
11051   rlevlrakwa  dmrsaegavd  ggadvdieea  sdddmfalld  delgln
```

SEQ ID NO. 8

NysE


```
  1    mttsteeslw  arcfhpapaa  pvrlfcfpha  ggsasfyfpv
 41    saqlssvaev  faiqypgrqd  rrkeagvsdl  atladqvyda
 81    lrpllkerps  tffghsmgat  lafevarrfe  addgdlvrlf
121    asgrrapsrv  reeavhrrsd  dgiveelkll  agtntallgd
161    eeilrmilpa  irsdyqaiet  yrcppdvtvr  apltvltgdr
201    dpktsldeae  awrghttgdf  dlkvlpgghf  fvsseapaii
241    dllrahlagn  g
```


SEQ ID NO. 9

NysR1


```
  1    mrkqsgssgl  lttlvgrdde  lrtlarhaaa  ardgraglvl  lhgpagmgkt
 51    sllrsftasd  vcrgmtvlyg  tcgetvagag  yggvrellgg  lglsggdarr
101    splleglaar  alpaltadpa  gpdaatgayp  vlhglywlaa  rlmaqrplvl
151    vlddvhwcde  rslawidfll  rraedlpllv  vlawrseaep  vapavladia
201    aqrrptvlgl  hplgpddige  mvrrvfrtta  apsfvsrvaa  vsggnplala
251    rlldelraeg  vrpdaagerr  aaevgshvla  rsvrcllerr  ppwvrgvara
301    iavlgpecte  llaalagvpa  atvdeallvl  rragilaadr  vdfvhdvvrs
351    avlddvappt  laelrtnaal  llsdagrpse  elagqlmllp  vldqpwmaav
401    lrdaaaqaes  rgapeagvrc  lyrvlevepd  nvavriqmar  alaeinppea
451    mrllkealsl  agdvrtraqv  avqygftcla  vqespsgvrm  ledalaelta
501    elgpepgpvd  relrtlvesv  llivgadekv  tigavrdraa  rltmppgdtp
551    aqrqmlamtt  vltamdgrda  rsavdqarra  lrapgvelep  wsllsasfal
601    sladevadaq  yaldlmlqyg  qdnaavwtyv  lalstrallh  hgvgafpeal
651    adaqtaveil  geerwadgav  lprvalatal  vdrgeperae  hvldgitrpr
701    lerfvieyhw  ylqarayarw  vrgdfqgald  lllacgrsle  esrfsnpafv
751    pwwadgavll  atldrhdqar  elaaygsela  erwgtarglg  lafmaqgvaa
```

```
801  pgragidhlt eavslladsp aramearael llghahlkrd dlraarehlr
851  aaadlaqrcg avklgvdark llvtaggrvr rmtaspldml tgmertvadl
901  avtgasnrai aealfvtvrt iethltsvyr klgvggrael savletrtat
951  sgrqppawvs qargra
```

5

SEO ID NO. 10

NysR2


```
10      1  vprskarnqp ttctpqcapd ahgdptmlle cgreqrligd llhrlgqgrp
       51  svlsltgrpg haqnalvrwg acrarhdglr vlraqatpae relrygavlq
      101  llavldgphg stldaairhd gpplpvpgi eevlrrtgta ptlvvvedvq
      151  wldpasltwl qillrhlgpd tplavlassc gdttafdtdp kapavpgppd
      201  tvpvarfvvp altdrgvaat vravcgtpgd eefiaaltsa tagnpailrd
15    251  alrafvdhgl padadhlpel haltagvvgd htvraldglp aevnavlral
      301  avcgdlldfh rvralagahs lsedrirtll asvgltvsvg dkvhirfpas
      351  karviedmpa aeradlyvra aelthscgvn dedvahlllr ssplgapwvv
      401  pllrrgfaaa lrredhhrac aclsralqep ldprerslLt lelaaaeava
      451  rpeagdrrlg elvrstvadt dptssgeqvg vraidlgfar gnsewvrrta
20    501  gealpyagpa dreelvalfw laavrdddap mipvvprlpd rpvppaqaga
      551  rawqlatage dadkarklar ialtggvnes lmmpklaaca alfatddnde
      601  avhgldtmlt aarsahlrsm aarifnlrar ihlcaarlea aerdldsaer
      651  alpptswhpr alpnliatri lvsmetgrpd rarrlaeapv paggeeqvww
      701  palllararv aaddgdweea lrlsrecgrw lfrrhwanpa mlswrplaae
25    751  aclklgdvte arrlrdeelf fadrwgtasa rgiarlttrr lfdddgdrav
      801  rrireaaall rdsparlayl wsrlsqagae tahgdtaaaa rswqavarmt
      851  aahpasrlat aartltvpsv pvatapptav vppgwrdlse aekdtvllaa
      901  rghgnrqiae qlavsrrtve lrlsnayrkl riggrkelyl llealegpva
      951  das
```

30

SEO ID NO. 11

NysR3


```
35      1  mllerenela riraaldaae agdssLLLin gplgsgrsal lrripelagd
       51  gtrvlrasaa wrerdfpfgi arqlfdhlls gaggagpaer tagaehfsrl
      101  mdtgdrptgt gpalevsqav lqgaqallad asaerrllil vddlqwadgp
```

```
151  slrwlahltr  rlhglrallv  ctladgdhrg  ryplvrevag  aahtvlrlap
201  lsrdatrvll  agpqgrppqd  alvravyeas  rgnplfltaf  rsalratgrp
251  pggdhfgavr  elsptvlrdr  laghlriqpq  pvrevavava  algdhsdpvl
301  laqlagvdei  gfagarralv  dagllargrd  vrfvhgvvrd  avdslltlde
351  rershddaad  llyrcgrpae  qvaghllavv  hpgrpwseav  lrsaahnalr
401  agrpadaary  lrrallhhrt  qdgcrarilv  dlataerald  pdacvrhvsq
451  avalldtsrd  raaavlripp  sllaapspsa  velvrqaaag  ldepgqrdee
501  gadelalrle  awlrhsghen  pvelassvar  lrrmgarppv  dsvaerelva
551  vllsagalsg  rlsaaeiadt  gnrilerepa  taahahtplp  lvmlslfvae
601  svqgvaswla  seqhtrrrya  tgaddvllta  erafvlvtqg  rpaaarehve
651  ralvmdagdw  sepavmmfaa  vafelrdpal  serilerird  rrpaglalta
701  tgqmlqaavd  vhfgrgrdal  dtllacgrrl  etvgwrnsal  lpwrpyaigl
751  hqrlgetdaa  lqlaedelrw  arewgattnl  gralrlkgwl  lqdegldllr
801  esveilrass  yatelartlv  vlgrrlpggp  eaeavlreaa  giaaacgvpw
851  laeraelglg  saivppvatl  tpserrvasl  vsrgltnqai  atelgvssra
901  vekhltsayr  klgvsgrrel  vnalpgr
```

SEQ ID NO. 12

NysR4

```
  1  visaqtapag  esvgpglmas  ldrdltikha  nqefrrrfdd  sagdvcgrsf
 51  rdlmhpsvqq  plmrqfsrli  egkrhrfash  vvavgaqdaa  fagtltasav
101  tgktpdiagi  lvlmdssgaa  daadagvvts  qkkflteida  rilegiaagl
151  stiplasrly  lsrqgveyhv  tgllrklrvp  nraalvsray  smgilnvgtw
201  ppkvvddfik
```

SEQ ID NO. 13

NysR5

```
  1  vdaegrrrdm  lelirrsgsa  dvvrlaeefa  vsketvrrdl  nvleghglir
 51  rrhggaypmv  rpgseavfvs  rtaqpipees  riataaaell  seaetvfide
101  gftpqliada  lprdrpltiv  taslpvvsaf  atspqanvll  lggrvrrgtt
151  atvdhwavhm  lsgfvidlaf  lgaegisrry  glttpdpava  evkaqairva
201  rrpvlagvht  kfgtasfcrf  gevgdletiv  tgaglpvaea  hryhlmgpkv
251  lrv
```

SEQ ID NO. 14

ORF2

```
  1  vaqdsgqtpr  sldhvdqalv  halqitpras  wtrigsvlgl  davtvarrwn
 51  rlvetgaawi  schpapvlaa  sgqgclafve  idcapgrlld  varalaaaph
101  vvalshvtgd  rdlqlnvmar  dpamlsrwvt  hdlaaldgvr  aarthlagpv
151  htegsrwrlr  algrhqvarl  aadasrhrtd  tpafvldeld  qqlvtalsvd
201  gratyralae  qcgagpdtvr  rrvqrlfaad  mlharcevar  plsewpvtvs
251  fwgqvpaarl.revtrrvtgm  revrlcasvi  srhnlhlvaw  vrslddaqrf
301  evrlaeraad  ltvteraval  whmkhgghll  deegyrvgvt  plalwreptd
351  arrg
```

SEQ ID NO. 15

ORF1

```
  1  mqpelerlrr  slhrepelgl  alprtqekvl  aaldglplei  tpgtgltsvt
 51  avlrggrpgg  avllrgdmda  lpvteesgvp  yaseipgrmh  acghdlhtag
101  lvgaarllae  rrerlhgdvv  fmfqpgeegh  ngagamieeg  vldaagkpld
151  aayalhvasn  qvpggvvitr  sgtitsasdl  ltvtvrgegg  hgstpatakd
201  pvpaacemvt  avqnwvtraf  difdpvvvtv  gtfhagtkas  vipdnrrvpg
251  hgaqlfrirp  gparrrelpr  lvraiaaahg  leaevdylrh  ypvtvndtde
301  tafavatard  lfapgevwes  piptngsedf  ayvlrrvpga  mllvgaapeg
351  sdwqhapmnh  spravfddrv  lyrqaallte  laarrlaata  paepavag
```

Translation products of SEQ ID No. 2 (Nys 2):

SEQ ID NO. 16

NysF

```
  1  vielilpatv  ateaayddrp  rpgdrllsse  reviaraves  rqrefttvrh
 51  larralrrlg  hpdrailpnr  rgapqwppgi  vgsmthcagy  raaavspael
101  saavsidaep  ngplpagvln  aialpserph  lvalaahrpd  vhwdrllfsa
151  kesvfkawyp  ltqreldfse  aeividptqg  aftarllvpg  pllggrrvtv
201  fpgrwhstpa  llttavhlpa  ptprrdrehr  thltvnsplp  rptfg
```

SEQ ID NO. 17

NysG

```
  1    maspddleee rtaprparrl vgllrphrrs valavsmgvg givlnafgpl
 51    llgrvtdlia dgvlggvpgp apgidfaaig rlllvllaly vvaslfmlaq
101    grlvasavwr tihelrrdar ekltrlplrh fdrqpagell srttndidnl
151    qqtlqqtlae litsifsllt mlvlmlvisp slavvmllsv pvsaliaari
201    skraqphyaa qwsangtlna hveevctgha likgfdrraa aeerfdacnd
251    avyraaakaq fasgamepvm mfvanlgyvl vavigawkvi ngtltlgdvq
301    afilyarqfs qpiveiasva grlqsgiasa qrvftlldap eqapdplrpg
351    tparaegrve ftdvsfrysp dtplienlsl tvepgstvai vgptgagktt
401    lgnllmrfye pdsgrilldg tdtatmtrdd lrsrfglvlq dtwlfggtia
451    eniaygapga cradieeaar atcadrfirt lpqgydtvld desgtvsage
501    kqlltvaraf larpavlvld eatssvdtrt evliqramns lragrtsfvi
551    ahrlstirda dlivvmdagr iveqgthdql lcaqglyarl haarthtpta
601    gaaag
```

SEQ ID NO. 18

NysH

```
  1    vllrllraql rpyawataal valqlvqilg tlllptlgaa lidqgvvrgd
 51    ggritelgvv mgvvalvqia aalgaaalaa rtatamgrdl rsalfrrild
101    fsareigrfg tpslltrsvn dvqqvqnlaq tgfgivvcap lmclgsvlla
151    lrqdvplall lvalvlvvav cfglllarmg tlyarmqltl drlgrllrea
201    itgvrvvrsf vrddherarf aqtndaflvv srrvgrliat mlpvvlllmn
251    gftvallwtg shridagrmp igslsallsy lslilmsvvm lafvflsvpr
301    arvcagriae vldtgssvap paapqpvrgp agrielcaag yrypgaeepv
351    lrdvdltvep geriavlgst gsgkttllnl vlrladateg avrvggtdvr
401    eltaatlaaa vgfvpqrpyl fsgtvasnlr fgrpdatdee lwealrvaqa
451    adfvarmpdg ldaeitqggg nvsggqrqrl slarallrrp eiylfddcfs
501    aldqatdaal rtalvpytag atvitvaqri sagrdadriv vldrgrvvaq
551    gthdvllrts ptyreialsq lteeeaahgl agrp
```

SEQ ID NO. 19

NysD3

```
5     1  mskralitgi tgqdgsylae hllsqgyqvw glirgqanpr ksrvsrlase
     51  ldfidgdlmd qgslvsavdt vqpdevynlg aisfvpmswq qaelvtevng
    101  mgvlrmleai rmvsglstsr tvsprgqirf yqasssemfg kaaetpqret
    151  tlfhprspyg aakayghyit rnyresfgmy avsgmlfnhe sprrgqefvt
    201  rkislavari kqglqdklal gnldavrdwg yagdyvramh lmlqqdagdd
10  251  yvigtgqmhs vrdavriafe hvglnwedyv vidpdlvrpa evevlcadsa
    301  kaqdrlgwkp dvdfptlmrm mvdsdlaqvs renqygdvll aanw
```

SEQ ID NO. 20

NysI aa sequence (partial)

```
15    1  mdneqklrdy lklatadlrr trrrvhkles aaqepvaiig mtcrypggvr
     51  spedlwrmve agehgvtpfp tdrgwdleal aaaptasggf lhdapdfdad
    101  ffgispreav amdpqqrvvl esaweafera gidptsvkgs rtgvfigama
20  151  qdyrvgpadg aegfqltgnt gsvlsgrisy tfgtvgpavt vdtacsssslv
    201  avhlatqalr agectlalag gvtimsgpgt fiemgrqggl sadgrcrsfg
    251  dtadgtgwae gvgilvlerl sdavrnghei lavvrgtavn qdgasnglta
    301  pngpsqqqvi qqalvnarla agdidvveah gtgttlgdpv eaqallatyg
    351  qnrpadrpll lgsvksnlsh tqaaagvagv ikmvmamrhg tlprtlhaee
25  401  pthhvdwsqg avrlltdttd wpatgaprra avssfgisgt nahtiieqap
    451  epqpedaata qddaagstpa tapvvpgvvp vllsgrtpda lrgqaaalra
    501  aldtgrrpdl ldlahslatt ragfehravl latdhpaltd gltaladadd
    551  paaapawitg ttraetrlav lftgqgaqrl gagrelaarf pafataldaa
    601  ldaftphldr plrevlwgtd aalldrtaya qpalfaveva lyrliesfgv
30  651  rpdhlaghsv geivaahlag vlsladaatl vaargrlmqa lpdggamiav
    701  qaseadvapl laghedqvai aavngpsavv lsgaeatvta laeqlaadgr
    751  ktrrlrvsha fhsplmepml dafravvedl tlqppllpvv snltgkpatv
    801  aqltsadywv dhvrhavrfa dgidwlarhd ttaflelgpd gvlsamaqdc
    851  ldaadadavt lpalragrpe ehtlttalag lhvhgatldw tgcfagtgar
35  901  rtdlptyafq rrrywpkalq sgtadlrsvg lgaahhplls aavsladagg
    951  tlltgrlsrq thpwladhtv rgttllpgta flelavragd evgcdrveel
```

```
        1001  tlaaplllpe  qggvqvqlwi  gnpdvsgrrt  vnvharpdtg  ddtpwtahat
        1051  gvlttadasr  qlpasseqgg  tplagdphpa  ldaaqwppag  aeplpldghy
        1101  drladggfgy  gpvfqglraa  wrggdvvyae  velpeagrsd  aeafglhpal
        1151  ldaalhaapf  tglgergrgg  lpfswegvsl  haggattlrv  rltpvaddal
   5    1201  altvadgtga  pvlsvdslvl  rsvatqqldt  aaavardalf  rldwtpvqpt
        1251  atdpgpvall  gadpfgllth  agfadapayp  dlaalaaadg  pvpttvvlsl
        1301  agtgddaadp  arsahrcaae  alaavqtwld  hherfaaarl  vfvtrgatvg
        1351  rdvaaaavwg  lvrsaqsenp  gcfalvdldp  dgavgaaalv  aalvsgepql
        1401  avrgdvlrva  rlvrrpltev  gagadgtgdg  vgdgsgvsfs  gegavlvtgg
  10    1451  tgglgavlar  hlvaeygvrd  lllvsrsger  avgagelvae  lagvgarvrv
        1501  vacdvtdraa  vvelvgghav  savvhaagvl  ddgmvgaltg  erlsavlrpk
        1551  vdavwhlhea  trgldldafv  vfsslagvfg  spgqanyaaa  nafldalmtr
        1601  rraeglpgls  lawgpweqsg  gmtgtltdvd  aerlarsgvp  plsvaqglal
        1651  fdaavagtda  tcvpvrldlp  vlrargevpp  llrslirvra  rraavagsat
  15    1701  agnlaqrlrr  ldedgrdemv  ldlvrgqval  vlghatggdv  dagrafrdlg
        1751  fdsltavelr  nrlntvtglr  lpatlvfdyp  tvrhlatyvl  dellgtdaev
        1801  atvqpaavav  addpivivgm  acrypggvss  pedlwrvlte  gtdavsgfpt
        1851  nrgwdvesly  hpdpdhpgts  ytrsggflhe  agefdpgffg  msprealatd
        1901  sqqrllless  weaieragid  pvslrgsrtg  vfagvmysdy  samlaspefe
  20    1951  gfqgsgssps  lasgrvaytl  glegpavtvd  tacssslvam  hwamqalrsg
        2001  ecglalaggv  tvmstpavfv  dfarqrglsp  dgrckafada  adgvgwsegv
        2051  gvlvlerqsd  avrngheila  vvrgsavnqd  gasngltapn  gpsqqrvirq
        2101  alasggltag  dvdvveahgt  gttlgdpiea  qallatygrd  reperplllg
        2151  svksnlghtq  aaagvagvik  mvlamrhgvv  prtlhvdaps  shvdwsegav
  25    2201  ellseqaawp  etgrvrragv  ssfgisgtnv  hviveqapga  kaiaaagaar
        2251  rtpgavpvll  sgrgrsalrg  qaarllghlq  arpdaelvdv  alslattrsr
        2301  feqraavvaq  drdqliaslg  alaadrpdpa  vvegeaagrg  rtavlftgqg
        2351  sqraamgrel  hevqpefaaa  fdavcavfdp  lldrplrevv  faedgsdeaa
        2401  lldetgwtqp  alfavevalf  rlveswgvrp  dfvaghsige  iaaahvagvl
  30    2451  tledacrlva  aratlmqalp  tggamiaiqa  tedeiaahld  dtvaiaavng
        2501  pqsvvisgde  eaaetiaatf  aergrktkrl  rvshafhspr  mdgmldafri
        2551  vaegltyrap  riplvsdltg  rraddaevct  aeywvrhvre  avrfadcvrt
        2601  lrdagattfl  elgsdgllta  maedtlgddh  daelvpmlra  graeelaaat
        2651  alarlqvrgv  dvdwaaylag  tgarrtdlpt  yafqhayywp  qlptpaaala
  35    2701  aadpadqqlw  aavergdare  ladilglgeq  dltpldsllp  altswrrgnq
        2751  ekhlldtlry  rvewtrlskp  tapvldgtwl  lvasdataad  qpalldglad
        2801  algshgarvr  rlllddscad  ravlaerlar  tadvdaatqv  lsvlplderd
```

```
     2851  addcppltrg  laltvalvqa  ladtgaqgrl  wtatrgavst  npadpvthpv
     2901  qaaawglgrg  valehprlwg  glvdlpqvfd  eragqrlagi  lavkdapdge
     2951  dqvalratgv  sgrrlvrhtv  ealptaaeft  atgtvlitgg  tgglgaevar
     3001  wlaragaqhl  vltsrrgpda  pgaaelrael  egygpsvsvv  acdvadrdal
   5 3051  aavltalpee  lpltgvvhta  gvghygpldt  lstaefaglt  aaklagaahl
     3101  dalladreld  ffvlfgsiag  vwgsgnqsay  gaanayldal  alhrrargla
     3151  atsvawgpwa  eagmaaddav  setlrrqglg  lldpapamte  lrravvrqdv
     3201  tvtvadvdwq  ryaplftsar  psaliaglpe  vralaadert  eqdatgasev
     3251  vtrvralaep  eqlrlltdlv  rtesatvlgh  ssadavpegr  afrdvgfdsl
  10 3301  tavelrkrlg  aatglslpst  mvfdyptple  laqylraeil  gavlevagpv
     3351  atggaddepi  aiigmacrfp  ggvsspeqlw  dlvasgtdai  sefpvnrgwq
     3401  tghlfdpdpd  rpgttystqg  gflheadefd  ptffgispre  alvmdpqqrl
     3451  llettwesfe  ragirpetlr  stltgtfvgs  syqeyglgag  dgteghmvtg
     3501  sspsvlsgrl  syvfglegpa  vtvdtacsss  lvalhlacqs  lrngesnlav
  15 3551  aggatimttp  npfiafsrqr  alakdgrcka  fsddadgmtl  aegvgvvlve
     3601  rlsdaqrngh  pvlavlrgsa  inqdgasngl  tapngpsqqr  virqalanar
     3651  lapgdidale  ahgtgtplgd  pieaqalfat  ygrdrdpesa  lllgsvksni
     3701  ghtqsaagia  svikmvmalr  hselpptlha  dapsshvdws  agtvrlltqa
     3751  rawpetgrpr  raavssfgis  gtnahvlleq  apvadtpaee  rpavapvpia
  20 3801  agvvpwvvta  rsaaalrgqa  erllahaetv  gtalpaagpl  diglslvsar
     3851  arfehravvv  ppagtdplaa  lravatdgps  pvvargvadv  egrtvfvfpg
     3901  qgsqwvgmgs  qlldesavfa  eriaecaaal  aeftdwslvd  vlrgvvgaps
     3951  lervdvvqpa  sfavmvslaa  lwrsrgvlpd  avvghsqgei  aaavvsgals
     4001  lrdgarvval  rsqaigrala  grggmmsval  svdvleprlv  efegrvsvaa
  25 4051  vngprsvvva  gepealdalh  arltaddira  rriavdyash  shqvedlhee
     4101  llevlaelap  rtsevpffst  vtgdwldtar  mdagywfrnl  rgrvrfadav
     4151  adllaaeyra  fvevsshpvl  smavqeaide  agvpavaagt  lrrdqggtdr
     4201  fllsaaevfv  rgvdvdwagl  fegtgasrid  lptyafqheh  lwavppapea
     4251  vaaadpddaa  fwtavedgdv  saltaalgtd  edsvaavlpa  ltswrrarrd
  30 4301  rstvdawryr  vawkplggtl  phpsltgtwl  lvtadgiddt  dvagaletyg
     4351  aevrrlvlde  ecvdravlre  rlagaedvtg  ivsvlaaaer  tdavpgtslv
     4401  lgtaltvali  qalgdaeida  pvwaltrgav  stgradelta  pvqaqvtgig
     4451  wtaalehpqr  wggtldlpaa  ldaraaqrla  avlsgalgsd  dqlairpsgv
     4501  ftrrivraea  tagrpagtwt  prgttlvtgg  sgtlaphlar  wlaqrgaehl
  35 4551  vlisrrgtaa  pgaaelvael  aesgteatva  acditdrdav  aalladlkad
     4601  grtvrtvvht  aatielhtld  attladfdrv  lhakvtgaqv  laellddeel
     4651  ddfvlyssta  gmwgsgahaa  yvagnaylaa  laehrrangl  palslswgiw
```

SUBSTITUTE SHEET (RULE 26)

```
        4701   addlklgrvd  pqmirrsgle  fmdpqlalsg  lqralddnen  vlavadvdwe
        4751   tyhpvytsgr  ptplfdevpe  vrrltaaaeq  sagtvaegef  aaalralsda
        4801   eqqrtlletv  rteaasvlgl  ssaedltdqr  afrdvgfdsl  tavglrnrla
        4851   svtgltlpst  mvfdypnpaa  laaylhgela  garsaaagaa  avptgapdad
   5    4901   dpiaivgmsc  rypggvgsae  dlwrialdev  daisgfpadr  gwdaeglydp
        4951   dpdrpgrtys  vqggflrdva  efdpgffgis  prealsmdpq  qrllletawe
        5001   afehagidpv  gqrgsrtgtf  vgasyqdyas  gvpnsegseg  hmitgtlssv
        5051   lsgrvsylfg  fegpavtldt  acssslvamh  lacqslrnge  sslalaggvs
        5101   imstpmsfvg  fsrqralaed  grckayadga  dgmtlaegvg  lvllerlsda
   10   5151   ranghqvlav  irgsavnqdg  asngltapng  psqqrvirqa  lansavapgd
        5201   idvleghgtg  talgdpieaq  allatygqdr  aperplllgs  vksnightqm
        5251   asgvasvikl  vralqegvvp  kslhidrpst  hvdwssgaig  lltertpwpe
        5301   tgrprraavs  sfgisgtnvh  tileqapade  aptpadpprd  glvpvllsgr
        5351   geaalraqaa  rllafveerp  eahltdlahs  latsraaler  raaviaadrd
   15   5401   tltrglrals  dgrpdpglvq  gtagrgrtaf  lftgqgsqrp  gmgrelhdry
        5451   pvfadaldev  larlddgpdr  plrevlfaap  dsaeaalldr  tgyaqpalfa
        5501   vevalfrllt  swgltpdyla  ghsvgelaaa  hvagvlsldd  actlvaargr
        5551   lmqalpegga  mvaleaaede  vlpllegltd  rvsvaavngp  rsvvvagvee
        5601   dvllladlfa  adgrrtkrlr  vshafhsplm  damlddfaav  argltyhppt
   20   5651   ipfvsnvsgg  lataeqvrtp  dywvghvraa  vrfadgidwl  atqgdvhtfl
        5701   elgpdgvlsa  maresltdps  rtallptlrg  drpeepalvt  avaaahahga
        5751   rvdwsgyfad  hgarrttlpt  yafqrerywp  dttaatsaht  pgsaldaefw
        5801   aaverddvaa  laasldldda  tvtamvpalt  awrrrrgeqt  eldswryrvt
        5851   wkprggatap  aaltgrwlvl  vphdhqdrqd  dataawaadv  etalgtttvr
   25   5901   ltvtttdraa  laariteaag  dqgpfsgvls  llplatgdag  hpgapaaltl
        5951   tttavqalgd  agidaplwnv  trgavavgra  eqvtapeqaa  vwglgraval
        6001   elparfggtl  dlpatldgqa  arrlravlaa  tdgedavalr  psgvflrrla
        6051   hapagpdtar  tafdpaagtv  litggtggig  ghvarrlard  gathllltsr
        6101   rgpaapgada  lraeleelga  rvtlaacdaa  drdalaalla  elpddaplca
   30   6151   vfhtagvved  hvvdaltpen  faavlraktv  aahhlhelta  dldlaafvlf
        6201   sstagvlgaa  gqgnyaaana  hldalaehrr  shgltalsva  wgpwagsgmv
        6251   adaaeltdrv  rrggfeplap  epavrallra  ienddttval  adidwerfqr
        6301   afaavrplpf  vadlpetgra  tpatatgaat  glrqqlaelp  eherpaavld
        6351   llrtqvaavl  ghadprtved  dhafrdlgfd  sltilelrna  lnaatglslp
   35   6401   atlvydlptp  remadfllae  llgtlptdta  atvastaspk  lsasfeqggt
        6451   pfddpiavig  igcrfpggvt  tpeelwqlld  egrdgisrfp  ddrgwdlaal
        6501   gagasdtleg  gfltgvadfd  arffgispre  alamdpqqrl  llettweale
```

```
6551  ragidpttlr  gsttgvfvgt  ngqdyptllr  rsasdvagyv  atgntasvms
6601  grlsyalgle  gpavtidtac  ssslvalhwa  gralragecd  lvvaggvsvm
6651  aspdsfvefs  tqgglapdgr  ckafsdaadg  tawsegvgil  vlerlsaarr
6701  nghqvlglir  gtavnqdgas  ngltapngls  qqrviaqala  darlrpadid
6751  aieahgtgtt  lgdpiearal  itaygrdrda  erplllgtvk  snightqaaa
6801  gaagvikmlm  amrhgtlprt  lhvgtpsshv  dwsggtvall  ddarpwprtg
6851  qprragvsaf  gvsgtnahvv  veqapeteap  aapaaepape  atptvvpwvv
6901  sgrsrealqa  qldrltahta  ahparsaadv  grslatdrtl  fphravllag
6951  pdgvreaara  aaprtpgrta  flfsgqgaqh  almghdlyqr  fpvyadaldt
7001  vlaqfdtvld  vplraalfaa  pgtpeaalld  qtgftqpalf  avevalfrla
7051  eswrltpdfv  aghsigei
```

SEO ID NO: 21

Glu(Asp)Leu Gly Phe(Leu, Val) Asp Ser Leu

SEO ID NO. 22

GAG/C CTG/C GGC/G T/CTG/C GAC TCC/G CTG/C

SEO ID NO. 23

Val Asp Thr Ala Cys Ser Ser

SEO ID NO. 24

G/CGA G/CGA G/ACA/ G/CGC C/GGT GTC G/CAC

SEO ID NO. 25

GTTGGTACCCCACTCCCGGTCCGCAC

SEO ID NO. 26

CCAGCCGCATGCACCACC

SEQ ID NO. 27

CCG CGT CGG ATC CGC CGA C

5   SEQ ID NO. 28

AGC CTT CGA ATT CGG CGC C

SEQ ID NO: 29

10

DNA sequence of ERD48.seq

```
    1   gatccgccga cgacaccggc cgccgcaccg tcaccgtcca cgcccgcccc
   51   gacgacaccg ccgaccgcac ctggacgctg cacgccaccg gtgtgctcgc
  101   caccacgcca ccggccgccg cggcgttcga caccacggtc tggccgcccg
  151   ccgacgccga acccctcacc accgacgact gctacgcaca cttcaccacc
  201   caccgcttcg cctacggccc cgccttccag ggcctgcggg ccgcctggcg
  251   cgccggcgac gtgctgtacg ccgaggtcgc cctgccggag tccgccaccg
  301   acgaagcggc cgccttcggc ctgcacccgg cgctcctgga cgccggcctg
  351   cacgccgcgc tcctcgccga cgaccgcgac accggactcc cgttctcctg
  401   ggaaggcgtc actctgcacg cctccggcgc caccgcgcta cgcgtccggc
  451   tcgccccgaa cggccccaac ggcctgtccg tcaccgccgc cgacccggcc
  501   ggcaaccccg tcgccaccgt cacccgcctg ctcgcccgcc ccctggacgc
  551   cgagcagttg accatccaca gcgccctgac ccgcgacgcg ctcttccacc
  601   tggactggac cccggtcccg cttcccgaca ccgccaactc cgcgccgccg
  651   gccctcctcg gccggacac cgccgtgctc gccgacgccc tcggcgaccc
  701   ggccgtcgca cgccacgcaa ccctcgacga cctcctggcc ggggacacca
  751   ccccgcccgc cacggtcctc gtccccctcg gcgccccact cgacggcgac
  801   accgcgcagc acgcgcacgc cctcacccgc agcgcgctga ccctcgtcca
  851   gcagtggctc gccaccgacc gcctcgccga ctcccgcctg gtcttcgtca
  901   cccacggagc cgtcgccacc gacgacgcgc cccccaccga cctggccgcc
  951   gccgcggtct ggggcctgat ccgctccgcg cagaccgaga accccggcac
 1001   cttcaccctc ctcgacctcg acaccgagcc cgactcgacc accgcgctca
 1051   gccgcgccct gaccctcgac gaaccacagc tcctcctccg cgccggccgc
 1101   gcccgcgccg cccgcctcac ccgcaccccc gccccacca ccaccaccca
 1151   cacgccgtgg tccgcggacg gaacggtgtt ggtgacgggt ggtacgggtg
 1201   gtctgggtgg gttggtggcc cggcatctgg tgcggtcgtg tggggtgcgg
```

```
       1251  catttgttgt tgaccagtcg ttctggtgtg ggtgctgcgg gtgcggccgg
       1301  gttggtcgcg gagttggagt cgttgggcgc gcgggttgtg gttgcggcgt
       1351  gtgatgtggg tgatggctcg gctgttgcgg agttggttgc cggtgtgtcg
       1401  gagtcgtatc cgttgtctgc ggtggtgcat gcggctggtg tgttggatga
    5  1451  cggtgtggtg ggttcgttga cgccggagcg gttggctgcg gtgttgcgtc
       1501  cgaaggtgga tggtgcgtgg aacctgcatg aggcgacgcg tggtctggat
       1551  ctggacgcgt ttgttgtctt ctcgtctgtt gcgggtgtgt tcggggtgc
       1601  gggtcaggcc aactatgcgg cgggtaatgc gtttttggac gcgttgatgg
       1651  ttcatcgggt ggctggtggg ttgcctggtg tgtcgttggc gtggggtgct
   10  1701  tgggatcagg gtgtggggat gacggcgggg ctgacggagc gggatgtccg
       1751  tcgtgctgct gagtcgggta tgccgttgtt gacggttgat cagggtgtgg
       1801  cgttgttcga tgcggcgttg gcgacgggga gtgccgcgtt ggtgccggtc
       1851  cgtctggacc tggccgcact gcgcacccgg ggcgacatcg caccgctcct
       1901  ccgcggcctg gtcaaggcgc ccatccgccg cgcagccgcc accacacccg
   15  1951  gcgacaccgg actcgccgag cagctcaccc ggctccagcg cgccgagcga
       2001  cgggacaccc tcctcgcgct cgtccgcgac caggccgcga tggtcctcgg
       2051  ccacacctcg ggcgacggcg tcgacccgtc ccgcgccttc cgcgacctcg
       2101  gcttcgactc gctcaccgcg gtcgaactcc gcaaccgcat cggcgcggcc
       2151  accggcctgc ggctaccggc cacggccgtc ttcgactacc ccaccgccga
   20  2201  tgccctcgcc gcacacctgc tcaccgaact gctcggcccc gacgccgagt
       2251  cggaccccga cgagcccggc gaccccaccg cgggaccgac cgacgacccc
       2301  atcgtcatca tcggcatgag ctgccgcttc cccggcgaca tcggctcgcc
       2351  ggaggacctg tggcgcctgc tcggcgacgg cgccgacgtc gtcaccgact
       2401  tcccgaccaa ccgcggctgg gacctggaca acctctacga ccccgacccc
   25  2451  gcgcacgccg gcacctcgta cgcccgcacc ggcggtttcc tgcacgacgc
       2501  cgccgacttc gacgccgact cttcggcat gagcccccgc gaggccatgg
       2551  ccacggactc ccagcagcgc ctgctgttgg agtcctcgtg ggaggcgatc
       2601  gagcgggccg gcatcgaccc gctgaccctg cgcgacagcc gcaccggcgt
       2651  cttcgccggc gtcatgtaca gcggctacgg cacccgcctc gacggcgccg
```

30

SEQ ID NO: 30

Translation product of SEQ ID NO. 29 (ERD48.seq)

```
35    1  saddtgrrtv tvharpddta drtwtlhatg vlattppaaa afdttvwppa
      51  daeplttddc yahftthrfa ygpafqglra awragdvlya evalpesatd
     101  eaaafglhpa lldaglhaal laddrdtglp fswegvtlha sgatalrvrl
```

```
      151  apngpnglsv  taadpagnpv  atvtrllarp  ldaeqltihs  altrdalfhl
      201  dwtpvplpdt  ansappallg  pdtavladal  gdpavarhat  lddllagdtt
      251  ppatvlvplg  apldgdtaqh  ahaltrsalt  lvqqwlatdr  ladsrlvfvt
      301  hgavatddap  ptdlaaaavw  glirsaqten  pgtftllddld  tepdsttals
  5   351  raltldepql  llragraraa  rltrtpaptt  tthtpwsadg  tvlvtggtgg
      401  lgglvarhlv  rscgvrhlll  tsrsgvgaag  aaglvaeles  lgarvvvaac
      451  dvgdgsavae  lvagvsesyp  lsavvhaagv  lddgvvgslt  perlaavlrp
      501  kvdgawnlhe  atrgldldaf  vvfssvagvf  ggagqanyaa  gnafldalmv
      551  hrvagglpgv  slawgawdqg  vgmtaglter  dvrraaesgm  plltvdqgva
 10   601  lfdaalatgs  aalvpvrldl  aalrtrgdia  pllrglvkap  irraaattpg
      651  dtglaeqltr  lqraerrdtl  lalvrdqaam  vlghtsgdgv  dpsrafrdlg
      701  fdsltavelr  nrigaatglr  lpatavfdyp  tadalaahll  tellgpdaes
      751  dpdepgdpta  gptddpivii  gmscrfpgdi  gspedlwrll  gdgadvvtdf
      801  ptnrgwdldn  lydpdpahag  tsyartggfl  hdaadfdadf  fgmspreama
 15   851  tdsqqrllle  ssweaierag  idpltlrdsr  tgvfagvmys  gygtrldga
```

SEQ ID NO. 31

CGCCGCATGCTGTTCTCACCCCACGT

SEQ ID NO. 32

GGCGCGACCCGGTTCGGCCT

SEQ ID NO. 33

GCGAGCGGCCGCTTCACCCCGCAACTCA

SEQ ID NO. 34

CGCGAAGCTTGGCCGACTGCTCGACGTC

SEQ ID NO. 35

DNA sequence - 125401 bps (entire cluster):


```
      1  gaattcggat  cgatgctgct  cgtggtgtgt  gccgcgggag  gtgccgacgg
```

```
   51  tgccgtgcgt gaatcggggc gtggtcgccg tcgcgatcgc ggtggtgtcg
  101  gtgctgctcg tgctgaccgt ggtggtggcg acggcctact tgccgtgagc
  151  ccggtccttg cgccgcgacc tgggcagcgg atgcacctct tcggtcgacg
  201  ccaaggggcg gtggcgcagg tcgagatagg cgcggacgga gacgccgggc
  251  tgccccggcg tctgccaggg cgcggagggg ccgctggaac gccacagacg
  301  gcccagctcg gccgcggcaa gggtggcggt ggcttcgtcg cggcggtga
  351  cgtcgatgac gaccatgccg ggttcgctga catgttcctc gtggatatgc
  401  acagttctgt atgacctatt tcgccccgtg gcgtaaggag tagccggcag
  451  gtttcatccg aaggtggggc gcgggagcgg cgagttgacc gtgaggtgcg
  501  tgcggtgttc ccggtcgcgc cggggggtgg gggccggcag gtggaccgcc
  551  gtggtcagca gggcgggggt gctgtgccag cggccgggga agaccgtgac
  601  ccggcgaccg cccagcaggg ggcccgggac caggagtcgg cgggtgaagg
  651  cgccctgtgt cgggtcgatg acgatctcgg cctcggagaa gtccagttcg
  701  cgctgggtca ggggatacca cgccttgaag acgctctcct tggcgctgaa
  751  gagcagccgg tcccagtgga cgtccggacg atgggccgcc agggccacga
  801  ggtgcggccg ttccgagggc agggcgatgg cgttcaggac gccggccggc
  851  agcggaccgt tcggttcggc gtcgatgctc accgcggccg acagctccgc
  901  gggggagacg gcggcggcgc ggtagccggc gcagtgcgtc atgctgccga
  951  cgatgcccgg cggccactgc ggggcgccgc gccgattggg cagtatggcc
 1001  cggtccgggt ggccgagccg gcgcagggcc cggcgggcga gatggcggac
 1051  ggtggtgaac tcgcgctgcc gggactccac ggcccgggcg atgacctcgc
 1101  gttccgagga gagcagccgg tcgccggggc gcggccggtc gtcgtacgcc
 1151  gcctcggtgg cgaccgtggc gggcaggatc agttcgatca ccgcatacct
 1201  ccggcgaacg gtaagaacag ggggttgctt cggggcacgg ttccgtcctt
 1251  gacggggcgc cgtgggcggc cgggtcagcc ggccgcggcg cccgcggtgg
 1301  gagtgtgggt gcgggctgcg tgcagccggg cgtacaggcc ctgcgcgcac
 1351  aggagttgat cgtgggtgcc ctgctcgacg atgcggccgg cgtccatcac
 1401  gacgatcagg tccgcgtcgc ggatcgtgga caggcggtgc gcgatcacga
 1451  agctggtccg gcccgcccgc agggagttca tggcgcgttg gatcaggacc
 1501  tcggtccggg tgtccacgga gctggtggcc tcgtccagca cgaggacggc
 1551  cggtctggcg aggaaggccc gggccacggt cagcagctgc ttctcgccgg
 1601  cgctgacggt gcccgactcg tcgtccagca cggtgtcgta gccctgcggc
 1651  agggtgcgga tgaagcggtc ggcacaggtc gcgcgggccg cctcctcgat
 1701  gtccgcacgg caggcaccgg gtgcgccgta cgcgatgttc tccgcgatgg
 1751  tgccgccgaa cagccaggtg tcctggagca ccagcccgaa gcgggaccgc
 1801  aggtcgtcgc gggtcatcgt cgcggtgtcg gtgccgtcca ggaggatgcg
 1851  gccggagtcc ggttcgtaga agcgcatcag gaggttgccg agggtggttt
```

```
1901   tgccggcgcc ggtggggccg acgatcgcca cggtgctgcc cggttccacg
1951   gtcagcgaga ggttctcgat gaggggcgtg tcgggggagt agcggaagga
2001   cacgtcggtg aactcgacgc ggccctcggc gcgggcgggc gtgccgggcc
2051   ggagcgggtc cggggcctgc tcgggggcgt cgagcagggt gaagacgcgc
2101   tgggcggagg cgatgccgga ctggagccgg cccgccaccg aggcgatctc
2151   cacgatcggc tggctgaact ggcgggcgta gaggatgaac gcctgcacgt
2201   caccgagggt cagggtgccg tttatgacct tccaggcgcc gatgacggcc
2251   accagcacat agccgaggtt ggcgacgaac atcatgaccg gttccatggc
2301   accggaggcg aactgcgcct tggccgcagc ccggtagacc gcgtcgttgc
2351   aggcgtcgaa gcgctcctcg ccgccgcgc gccggtcgaa gcccttgatc
2401   agcgcatgac cggtgcacac ctcctccaca tgggcgttga gggtgccgtt
2451   cgcggaccac tgcgcggcgt agtggggctg cgcgcgcttg ctgatccggg
2501   ccgcgatcaa cgccgagacc ggcacgctga gcagcatcac cacggccagc
2551   gacggcgaga tcaccagcat cagcaccagc atcgtcaaca gcgagaagat
2601   cgaggtgatc agctcggcga gggtctgctg gagggtctgt tggaggttgt
2651   cgatgtcgtt ggtggtgcgg ctgagcagct caccggccgg ctgccggtcg
2701   aagtggcgca gcggcagccg ggtcagcttc tcccgggcgt cgcggcgcag
2751   ttcgtggatg gtgcgccaca ccgcggacgc caccagccgg ccctgcgcca
2801   gcatgaacag cgacgccacg acgtagagcg ccagcaggac cagcagcagc
2851   cggccgatcg cggcgaagtc gatcccgggc gccgggcccg ggacgccgcc
2901   gagcacgccg tcggcgatca gatcggtgac ccggccgagc agcagcgggc
2951   cgaacgcgtt gagcacgatc ccgccgacgc ccatcgacac ggcgagtgcc
3001   acggagcggc ggtgcggacg cagcaggccg acgagccgac gtgccggccg
3051   gggcgcggtc cgctcctcct caaggtcgtc cggcgaggcc atgggcggct
3101   tcctcctcgg tcagctgcga gagcgcgatc tcgcggtagg tcgggctggt
3151   gcgcagcagc acgtcgtggg tgccctgggc gaccacgcgc ccgcggtcca
3201   gcaccacgat ccggtccgcg tcgcggccgg cggagatccg ctgcgccacg
3251   gtgatcaccg tggcgccgc ggtgtacggc accagcgcgg tccgcagcgc
3301   cgcatcggtg gcctggtcga gcgccgagaa acagtcgtcg aagagataga
3351   tctccggccg gcgcagcaac gcccgggcga gggacaggcg ttggcgctgg
3401   ccgccggaga cattgccgcc gccctgggtg atctccgcgt cgaggccgtc
3451   cggcatccgc gccacgaagt cggccgcctg ggcgaccgc agcgcctccc
3501   acagctcctc gtcggtggcg tccggccgcc cgaagcgcag attgctcgcc
3551   acggtgccgg agaacaggta cggccgctgc ggcacgaacc cgacggcggc
3601   ggcgagcgtg gccgcggtca gctcgcggac gtcggtgccg ccgacccgca
3651   ccgcgccctc ggtggcgtcg gccagccgca gcaccaagtt caacagggtc
3701   gtcttgccgc tgccggtgct gccgagcacg gcgatccgct cgcccggctc
```

```
        3751  gacggtcagg  tcgacgtccc  gcagcacggg  ctcctcggcg  cccgggtagc

        3801  ggtacccggc  cgcgcacagt  tcgatccggc  cggcgggccc  gcgcaccggc

        3851  tgcggcgcgg  ccggcggcgc  cacgctcgac  ccggtgtcca  ggacctccgc

        3901  gatccggccg  gcacagaccc  gggcccgcgg  caccgacagg  aacacgaagg

 5      3951  cgagcatcac  gacggacatc  aggatcagcg  agagatagct  caggagggcg

        4001  ctgagcgagc  cgatcggcat  ccggcccgcg  tcgatccggt  gggagccggt

        4051  ccacagcagg  gctacggtga  aaccgttcat  cagcagcagc  acgaccggca

        4101  gcatcgtcgc  gatcagccga  cccacccgcc  gcgacaccac  gaggaacgcg

        4151  tcgttggtct  gcgcgaaccg  cgcgcgctcg  tggtcgtcgc  ggacgaagga

 10     4201  ccggaccacc  cgcaccccgg  tgatcgcctc  gcgcagcagc  cgccccagcc

        4251  ggtccagggt  cagctgcatc  cgcgcgtaca  gggtgcccat  ccgggccagc

        4301  agcaggccga  agcagaccgc  caccaccagc  accagcgcca  ccagcagcag

        4351  tgccagcgga  acgtcctggc  gcagcgccag  cagcacgctg  cccaggcaca

        4401  tcagcggcgc  gcagacgacg  atgccgaagc  cggtctgggc  gaggttctgc

 15     4451  acctgctgca  cgtcgttcac  cgaccgggtc  agcagggagg  gggtgccgaa

        4501  ccggccgatc  tcgcgggcgg  agaagtccag  gatgcggcgg  aagagcgcgg

        4551  accgcagatc  gcggcccatc  gccgtcgcgg  tccgggcggc  cagcgcggcc

        4601  gcaccgagcg  cggcggcgat  ctgcaccagc  gccaccacgc  ccatcaccac

        4651  acccagctcg  gtgatgcgcc  cgccgtcgcc  gcgcaccacc  ccctggtcga

 20     4701  taagtgcggc  gcccagcgtc  ggcagcagca  aagtgcccag  gatctggacg

        4751  agttgaaggg  cgacgagagc  ggcggtggcc  caggcgtagg  ggcgcagctg

        4801  tgcccgcaga  agtctcaaca  gcacggagga  acaccccggt  tgacggcggc

        4851  gtgcggccgc  gcggtggacg  gagtcggccg  gcccgccgac  ccggtcattg

        4901  caccgcagtc  cgctccgaaa  tttcactagt  gttgggggtg  gcaacggact

 25     4951  tttgacggcc  cggtactcgt  gaattcccta  gaaagcccgg  gatgcgttga

        5001  cagcatcttc  cgcggcttgc  gagcgtgcgg  gtgtctatcg  tccggactgc

        5051  gtattcgacg  ccaggtcgtg  gccgagctca  agcgtttgga  cggtctcttt

        5101  cgtgttcgag  aagggtttgc  catgtccaaa  cgagcgctga  tcaccggaat

        5151  caccggccag  gacggctcct  atctcgcgga  gcacctgctg  tcccagggct

 30     5201  accaggtgtg  gggtctgatc  cgcggccagg  ccaatccccg  caagtcccgg

        5251  gtcagccgcc  tcgcctccga  actcgacttc  atcgacgggg  acctgatgga

        5301  ccagggcagc  ctggtctccg  ccgtcgacac  cgtgcagccc  gacgaggtct

        5351  acaacctcgg  cgccatctcg  ttcgtgccga  tgtcctggca  gcaggccgag

        5401  ctggtcaccg  aggtcaacgg  catgggcgtg  ctgcgcatgc  tggaagccat

 35     5451  ccgcatggtc  agcggactgt  ccacctcccg  cacggtcagc  ccgcgcggcc

        5501  agatccgctt  ctaccaggcg  tccagctcgg  agatgttcgg  caaggccgcc

        5551  gagacgccgc  agcgcgagac  caccctcttc  cacccgcgca  gccctacgg
```

```
       5601  cgcggcaaag gcgtacgggc actacatcac ccgcaactac cgcgagtcct
       5651  tcggcatgta cgcggtctcc ggcatgctct tcaaccacga atccccgcgc
       5701  cgcggccagg aattcgtcac ccgcaagatc agcctggcgg tcgcccgcat
       5751  caagcaaggc ctccaggaca agctggcact cggcaacctc gacgcggtgc
   5   5801  gcgactgggg ctatgccggc gactacgtcc gcgccatgca cctgatgctc
       5851  cagcaggacg ccggcgacga ctacgtcatc ggcaccgggc agatgcactc
       5901  ggtgcgcgac gcggttcgga tcgcgttcga acacgtcggc ctgaactggg
       5951  aggactacgt cgtcatcgac cccgacctgg tgcggcccgc cgaggtcgag
       6001  gtgctgtgcg ccgacagcgc caaggcccag gaccgcctcg gctggaagcc
  10   6051  ggacgtcgac ttccccaccc tcatgcgcat gatggtcgat tccgacctgg
       6101  cgcaggtttc ccgcgaaaac caatacggcg acgtgctgct cgccgccaac
       6151  tggtagcagt tctcaagctt tcgaaaacta gtgaattcct gccggaattc
       6201  cgacgacact cgccacatgg attcccccgg tgagtggcga atccaggtgg
       6251  cgaatccgaa cgtaccgccg aaacggcgtg gagaagtcgg acgccattca
  15   6301  cgtgcggccg tcggctgtcg agatgggttg agttgagatg gacaacgaac
       6351  aaaaactccg ggattacctc aagcttgcga cggccgacct tcgacgcacc
       6401  cggcggcgcg tccacaagct ggagtcggcg gcccaggaac cggtggccat
       6451  catcggcatg acctgtcgct accccggcgg cgtccgcagc cccgaagacc
       6501  tctggcgcat ggtcgaggcc ggcgagcacg gcgtcacccc gttccccacc
  20   6551  gaccgcggtt gggacctgga ggcgctggcc gccgcgccga ccgcctccgg
       6601  cggattcctg cacgacgcac ccgacttcga cgcggacttc ttcggcatct
       6651  cgccgcgcga ggcggtcgcc atggacccgc aacagcgcgt cgtcctggaa
       6701  tccgcctggg aggcgttcga acgcgccggc atcgacccga cgtccgtgaa
       6751  gggcagccgc accggagtct tcatcggcgc gatggcccag gactaccggg
  25   6801  tcggccccgc cgacggcgcc gagggcttcc aactcaccgg caacaccggc
       6851  agcgtgctgt ccggccgcat ctcctacacc ttcggcacgg tcggccccgc
       6901  cgtcaccgtc gacaccgcct gctcctcctc cctcgtcgcc gtccacctcg
       6951  ccacccaggc gctgcgggcc ggcgagtgca ccctcgccct cgccggcggc
       7001  gtcaccatca tgtccggccc cggcaccttc atcgaaatgg ccgccaggg
  30   7051  cgggctctcc gccgacggcc gctgccgctc cttcggcgac accgccgacg
       7101  gcaccggctg ggccgaaggc gtcggcatcc tcgtcctgga acggctgtcc
       7151  gacgccgtcc gcaacggcca cgagatcctc gccgtcgtcc gcggcaccgc
       7201  cgtcaaccag gacggcgcct ccaacggcct gaccgccccc aacggcccct
       7251  cccagcagca ggtcatccag caggccctgg tcaacgcccg actcgccgcc
  35   7301  ggggacatcg acgtcgtcga ggcgcacggc accggcacca ccctcggcga
       7351  ccccgtcgag gcccaggccc tgctcgccac ctacgggcag aaccgcccgg
       7401  cggaccggcc gctgctgctg ggctcggtca agtccaacct cagccacacc
```

```
        7451  caggccgccg  ccggcgtcgc  cggcgtgatc  aagatggtca  tggcgatgcg
        7501  gcacggcacc  ctgccgcgca  ccctgcacgc  cgaggagccc  acccaccacg
        7551  tcgactggtc  gcagggcgcc  gtgcggctgc  tgaccgacac  caccgactgg
        7601  cccgccaccg  gggcgccgcg  ccgcgccgcc  gtctcctcct  tcggcatcag
   5    7651  cggcaccaac  gcccacacca  tcatcgagca  ggcccccgaa  ccgcagcccg
        7701  aggacgccgc  gaccgcgcag  gacgacgccg  ccggcagcac  gccggccacc
        7751  gcccccgtag  tgcccggcgt  cgtaccggtc  ctgctctccg  gccgcacccc
        7801  ggacgccctg  cgcggccagg  ccgcggccct  gcgcgccgcc  ctcgacaccg
        7851  gccggcggcc  cgacctgctc  gacctcgcac  actccctcgc  caccacccgc
  10    7901  gccgggttcg  agcaccgcgc  cgtcctcctc  gccaccgacc  accccgccct
        7951  gaccgacggc  ctcaccgccc  tcgccgacgc  cgacgacccg  gccgccgccc
        8001  ccgcctggat  caccggcacc  acccgggccg  agacccggct  cgccgtcctg
        8051  ttcaccggcc  agggcgccca  acgcctcggc  gcgggacggg  aactcgccgc
        8101  ccgtttcccg  gcgttcgcca  ccgccctcga  cgcggcgctc  gacgccttca
  15    8151  ccccgcacct  cgaccgcccc  ctgcgcgagg  tcctgtgggg  caccgacgcc
        8201  gccctgctcg  accgcaccgc  atacgcccag  ccggccctct  tcgccgtcga
        8251  agtggcgctc  taccggctga  tcgaatcgtt  cggcgtccgc  cccgaccacc
        8301  tcgccggcca  ctccgtcggc  gagatcgtcg  ccgcgcacct  cgccggggtc
        8351  ctctccctgg  ccgacgccgc  caccctcgtc  gccgcccgcg  gtcgcctgat
  20    8401  gcaggcgctg  cccgacggcg  gggcgatgat  cgccgtccag  gcgtcggaag
        8451  ccgacgtcgc  cccgctgctc  gccgggcacg  aggaccaggt  cgcgatcgcc
        8501  gccgtcaacg  gcccctccgc  cgtcgtcctg  tccggcgccg  aagccaccgt
        8551  caccgcgctc  gccgaacagc  tcgccgccga  cggccgcaag  acccgccggc
        8601  tgcgcgtctc  gcacgccttc  cactcgccgc  tcatggagcc  gatgctcgac
  25    8651  gccttccgcg  ccgtcgtcga  agacctcacg  ctccagccgc  cgctcctgcc
        8701  ggtcgtctcc  aacctgaccg  gcaagcccgc  caccgtcgcc  caactcacct
        8751  ccgccgacta  ctgggtcgac  cacgtccggc  acgccgtccg  cttcgccgac
        8801  ggcatcgact  ggctcgcccg  gcacgacacc  accgccttcc  tcgaactcgg
        8851  ccccgacggc  gtgctgtccg  ccatggccca  ggactgcctg  gacgccgccg
  30    8901  acgcagacgc  cgtcaccctc  cccgccctgc  gcgccgggcg  ccccgaggag
        8951  cacaccctca  ccaccgccct  cgccggtctg  cacgtccacg  gcgccaccct
        9001  ggactggacc  ggctgcttcg  ccggcaccgg  cgcccgccgc  accgacctgc
        9051  cgacctacgc  cttccagcgc  cgccgctact  ggcccaaggc  cctccagagc
        9101  ggcaccgccg  acctgcgctc  ggtcggcctc  ggtgccgccc  accacccgct
  35    9151  gctctccgcc  gccgtctccc  tcgccgacgc  aggcggcacc  ctgctcaccg
        9201  gccgcctctc  ccggcagacc  cacccctggc  tcgccgacca  caccgtccgc
        9251  ggcaccaccc  tgctgcccgg  taccgccttc  ctcgaactcg  ccgtccgcgc
```

```
              9301  cggcgacgag gtcggctgcg accgcgtcga ggaactcacc ctcgccgcac
              9351  cgctcctgct gcccgaacag ggcggcgtcc aggtccagtt gtggatcggc
              9401  aaccccgacg tgtccggtcg ccgcaccgtc aacgtccacg cccgccccga
              9451  caccggcgac gacacccct ggaccgccca cgccaccggc gtcctcacca
         5    9501  ccgccgacgc ctcccgccag ctcccggctt cgtccgagca gggcggcacc
              9551  cccctcgccg gcgaccccca ccccgccctc gacgcggccc agtggccccc
              9601  ggccggcgcc gaaccgctgc cgctggacgg ccactacgac cgcctcgccg
              9651  acggcggctt cggctacggc ccggtcttcc agggcctgcg cgccgcctgg
              9701  cgcggcggcg acgtcgtcta cgccgaggtc gagctgcccg aggccggccg
         10   9751  gtccgacgcc gaggcgttcg gcctccaccc cgccctgctc gacgccgccc
              9801  tgcacgccgc gcccttcacc ggcctcggcg aacgcggccg gggcggcctg
              9851  ccgttctcct gggagggcgt ctccctccac gccggcggcg ccaccaccct
              9901  ccgcgtccgc ctgaccccgg tcgccgacga cgcgctcgcc ctgaccgtcg
              9951  ccgacggcac cggcgcgccc gtgctgtccg tcgactcgct cgtcctgcgc
         15   10001 agcgtggcga cccaacagct cgacacggcc gccgccgtcg cccgtgacgc
              10051 cctcttccgc ctcgactgga cccccgtcca gccgaccgcc accgaccccg
              10101 ggcccgtcgc cctcctcggc gccgacccct cggcctgct  cacccacgcc
              10151 ggattcgccg acgccccggc atacccggac ctcgccgccc tcgccgcggc
              10201 ggacggcccg gtcccgacca ccgtcgtgct gtccctcgcc ggcaccgggg
         20   10251 acgacgcggc cgacccggcc cggtccgcac accgctgcgc cgcggaggcc
              10301 ctcgccgccg tacagacctg gctcgaccac catgagcgct tcgccgccgc
              10351 ccgcctggtc ttcgtgaccc gtggtgcgac ggtcgggcgt gatgttgctg
              10401 ctgctgcggt gtggggtctg gtgcgttcgg cgcagtcgga gaatccgggg
              10451 tgttttgctc tggtcgatct ggatccggat ggtgcggtgg gtgcggctgc
         25   10501 gctcgtcgct gcgttggtca gtggtgagcc gcagcttgcg gtgcgcggtg
              10551 atgtgttgcg ggtcgcgcgt ctggtgcggc ggccgctcac cgaggtcggt
              10601 gcgggtgctg atggcaccgg ggatggcgtc ggggatggct ctggtgtgtc
              10651 gttctcgggt gagggtgcgg tcctggtcac tggtggtacg ggtggtctgg
              10701 gtgcggtgtt ggcgcgtcat ctggtggccg agtatggggt gcgggatctg
         30   10751 ctgttggtca gtcgcagtgg tgaacgtgcc gtgggtgctg gggagttggt
              10801 ggcggagctt gcgggtgtgg gtgcgcgggt gcgggtggtt gcgtgtgatg
              10851 tgaccgatcg tgccgcggtg gtggagttgg ttggcgggca tgcggtgtcc
              10901 gcggtggttc atgcggctgg tgtgctggat gacggcatgg tgggtgcgtt
              10951 gaccggggag cggttgtccg cggtgctgcg gccgaaggtg gatgctgtct
         35   11001 ggcatctaca tgaggcgacc cgcggcctgg acctggacgc gttcgtcgtc
              11051 ttctcctccc tcgccggggt cttcggcagt cccggccagg ccaactacgc
              11101 ggccgcgaac gccttcctgg acgcgctgat gacgcggcgc cgggcggagg
```

```
11151  gactgcccgg cctgtcactc gcatggggac cgtgggagca gtcgggcgga
11201  atgacgggca ccctgacgga cgtcgacgcc gaacggctgg cccgctccgg
11251  tgtcccgccg ctctccgtgg cgcagggcct ggccctcttc gacgctgccg
11301  tggccgggac cgacgccacc tgcgttccgg tccgcctgga cctccccgtc
11351  ctccgcgcac ggggtgaagt gccgccgctg ctgaggtcgt tgatccgcgt
11401  ccgggcgcgc cgagccgccg tcgccggctc cgccaccgcg ggcaacctcg
11451  cccagcgcct gcgccgcctg gacgaggacg gccgcgacga gatggtcctg
11501  gacctcgtcc gcggtcaggt cgccctcgtc ctcggccacg cgaccggtgg
11551  cgacgtcgac gccggccgtg ccttccgcga cctcggcttc gactcgctga
11601  ccgccgtcga actgcgcaac cgcctcaaca ccgtcaccgg cctgcgcctg
11651  cccgccaccc tggtcttcga ctacccgacc gtccggcacc tcgccacgta
11701  cgtcctggac gagttgttgg gcacggatgc cgaggtggcg accgtgcagc
11751  cggccgcggt tgcggtggcg gacgatccga tcgtcatcgt gggcatggcc
11801  tgccgctacc ccggtggcgt cagctccccc gaggacctgt ggcgcgtgct
11851  caccgaaggc accgacgccg tctcgggctt cccgaccaac cgtggttggg
11901  acgtcgaatc cctctatcac ccggaccctg accacccggg tacctcctac
11951  acgcgctcgg gtgggttcct gcatgaggcg ggggagttcg atccggggtt
12001  cttcgggatg agtccgcggg aggcgttggc gaccgattcc cagcagcggt
12051  tgttgttgga gtcgtcgtgg gaggcgatcg agcgggccgg tattgatccg
12101  gtgagtttgc ggggtagtcg gacgggtgtg ttcgcggggg tgatgtacag
12151  cgattacagc gcgatgttgg cgagtccgga gttcgagggt ttccagggca
12201  gtgggagttc gccgagtttg gcgtcgggtc gggtggccta cacgttgggg
12251  ttggaaggcc cggcggtgac ggtggatacg gcgtgttcgt cgtcgttggt
12301  ggcgatgcac tgggcgatgc aggcgttgcg tagtggtgag tgtgggttgg
12351  cgttggccgg tggtgtgacg gtgatgtcga cgcctgcggt gtttgtggac
12401  tttgctcggc agcggggttt gtcgccggat ggccggtgca aggcgtttgc
12451  ggatgcggcc gatggtgtgg gctggtccga gggcgtcggc gtgttggtcc
12501  tggagcggca gtcggacgcg gtgcgcaatg gtcacgagat tttggctgtg
12551  gtgcggggtt cggcggtcaa ccaggatggt gcgtccaatg gtttgacggc
12601  gcccaatggt ccgtcgcagc agcgggtgat ccggcaggcg ttggccagtg
12651  gtggtctgac ggccggtgac gtggatgtgg tggaggcgca tggtacgggt
12701  acgacgctcg gtgatccgat cgaggcgcag gcgttgttgg cgacgtatgg
12751  gcgggatcgt gagcctgagc ggccgttgtt gttgggttcg gtgaagtcga
12801  atctggggca tacgcaggct gctgcgggtg tggcgggcgt tatcaagatg
12851  gtgttggcga tgcggcatgg tgtggtgccg cggacgttgc atgtggatgc
12901  gccttcttcg catgtggact ggtccgaggg tgcggtggag ctgctcagtg
12951  agcaggcggc ctggccggag acgggtcggg tgcggcgggc gggtgtctcc
```

```
        13001   tccttcggca tcagcggtac caatgtgcat gtcatcgtcg agcaggcgcc
        13051   gggcgccaag gcgatcgccg cggccggtgc ggcgcggcgc acgccgggtg
        13101   ccgtgccggt gctgctctcg gggcgtggcc ggagtgccct gcggggccag
        13151   gccgcccgcc tgctcggaca cctccaggcc cgacccgacg ccgaactcgt
   5    13201   cgatgtcgca ctgtcgttgg cgaccacccg gtcccgcttc gagcagcggg
        13251   ccgccgtcgt ggcgcaggac cgcgaccagc tgatcgcctc gctggggggcg
        13301   ctggccgccg accgccccga ccccgccgtc gtcgagggcg aggccgccgg
        13351   acgcggccgg accgcggtgc tgttcaccgg acagggcagc cagcgggccg
        13401   ccatggggcg tgaactccac gaggtgcagc cggagttcgc cgcggcgttc
   10   13451   gacgcggtgt gtgccgtttt cgacccgctg ttggaccggc cgctgcgcga
        13501   ggtggtgttc gccgaggacg gcagcgacga ggccgcactg ctggacgaga
        13551   ccggttggac gcagccggct ctgttcgccg tcgaggtggc gctgttccgg
        13601   ctggtggaga gttggggtgt ccgtccggac ttcgtggccg gccattccat
        13651   cggtgagatc gcggcggcgc acgtcgccgg ggtgctgacg ttggaggacg
   15   13701   cctgccgtct ggtggccgcg cgggcgacgc tgatgcaggc gctgccgacc
        13751   ggcggcgcga tgatcgcgat ccaggccacc gaggacgaga tcgcggcgca
        13801   cctcgacgac acggtggcga tcgccgccgt caacgggccg cagtccgtgg
        13851   tgatctccgg tgacgaggag gccgccgaaa cgatcgccgc cacgttcgcc
        13901   gaacgcgggc gcaagaccaa gcggctgcgg gtgagccatg ccttccactc
   20   13951   gccgcggatg gacgggatgc tggacgcctt ccggatcgtc gccgaggggc
        14001   tgacctaccg ggcgccgcgc atcccgctcg tctccgacct caccggccgg
        14051   cgcgccgacg atgcggaggt gtgcaccgcg gagtactggg tgcggcacgt
        14101   ccgagaggcc gtgcggttcg ccgactgcgt gcggacgctg cgcgacgccg
        14151   gggccaccac cttcctggaa ctgggctccg acggcctgct gaccgcgatg
   25   14201   gccgaggaca ccctcggtga cgaccacgac gccgaactgg tgccgatgct
        14251   gcgcgccggg cgcgccgagg aactggccgc ggccaccgcc ctggcccgcc
        14301   tccaggtgcg cggcgtggac gtggactggg cggcgtacct cgccggcacc
        14351   ggcgcccgac gcaccgacct gccgacctac gccttccagc acgcgtacta
        14401   ctggccgcag ctgccgaccc cggccgccgc cctcgccgcc gccgatcccg
   30   14451   ccgaccagca gctgtgggcc gctgtggagc gcggcgacgc ccgcgaactc
        14501   gccgacatcc tcggcctggg cgaacaggac ctcacgccgc tggactccct
        14551   gctgcccgcc ctcacctcgt ggcggcgcgg caaccaggag aagcacctcc
        14601   tggacaccct gcgctaccgc gtggagtgga cacgactgag caagccgacc
        14651   gccccggtcc tcgacggcac ctggctgctg gtcgcctccg acgccaccgc
   35   14701   ggccgaccag ccagccctcc tcgacggcct ggccgacgcc ctcggctcgc
        14751   acggcgcgcg ggtgcgtcgc ctgcttctgg acgactcctg cgcggaccgc
        14801   gcggtgctcg ccgaacgact ggcgcggacc gccgacgtgg acgccgcgac
```

```
        14851  ccaggtgctg  tccgtgctgc  cgctcgacga  gcgggacgcc  gacgactgcc
        14901  cgccgctcac  ccgcggactg  gcgctgaccg  tcgcgctcgt  ccaggccctc
        14951  gccgacaccg  gcgcccaggg  ccggctgtgg  accgccaccc  gcggcgccgt
        15001  ctccaccaac  cccgccgacc  cggtcaccca  ccccgtccag  gccgctgcct
     5  15051  ggggcctggg  ccggggcgtc  gccctggagc  acccacggct  gtggggcggc
        15101  ctggtcgacc  tgccgcaggt  cttcgacgag  cgggccggac  agcggctcgc
        15151  cgggatcctc  gccgtcaagg  acgcaccgga  cggcgaggac  caggtggcgc
        15201  tgcgggccac  cggagtctcc  ggccgccggc  tcgtccgcca  caccgtcgaa
        15251  gcgctgccca  cggccgcgga  gttcaccgcc  accggcactg  tcctgatcac
    10  15301  tggtggcacc  ggtggcctgg  gcgccgaggt  cgcccggtgg  ctggcccgcg
        15351  ccggcgccca  gcacctcgtc  ctgaccagcc  gccgcggccc  ggacgcgccg
        15401  ggcgccgccg  aactccgggc  cgaactggag  ggctacgggc  cgtcggtgtc
        15451  cgtcgtcgcc  tgcgacgtcg  ccgaccggga  cgcgctcgcc  gccgtcctca
        15501  ccgcactgcc  cgaggaactg  ccgctgaccg  gtgtcgtgca  caccgcaggc
    15  15551  gtcggccact  acggcccgct  ggacaccctg  agcaccgccg  agttcgccgg
        15601  cctcaccgcc  gccaagctcg  ccggcgccgc  ccacctcgac  gccctgctcg
        15651  ccgaccgcga  actggacttc  ttcgtcctct  tcggctccat  cgccggtgtc
        15701  tggggcagtg  gcaaccagag  cgcctacggc  gccgccaacg  cctacctcga
        15751  cgcgctcgcc  ctgcaccgcc  gcgcccgcgg  cctcgccgcg  acctccgtcg
    20  15801  cctggggccc  gtgggccgag  gccggcatgg  ccgccgacga  tgccgtttcc
        15851  gagaccctgc  gccgccaggg  cctcggcctg  ctcgacccgg  ccccggccat
        15901  gaccgagctg  cgccgcgccg  tcgtccggca  ggacgtcacc  gtcaccgtcg
        15951  ccgacgtgga  ctggcagcgc  tacgcaccgc  tgttcacctc  cgccggccc
        16001  agcgccctga  tcgccggcct  gcccgaggtc  cgcgccctcg  ccgccgacga
    25  16051  gcgcaccgag  caggacgcca  ccggcgcctc  cgaggtcgtc  acccgcgtcc
        16101  gcgccctggc  cgaacccgag  caactgcgcc  tgctgaccga  cctcgtccgc
        16151  accgagtccg  ccaccgtcct  cggccacagc  tccgccgacg  ccgtgcccga
        16201  gggccgcgcc  ttccgcgacg  tcggcttcga  ctcgctgacc  gcggtcgagc
        16251  tccgcaagcg  cctgggcgcc  gcgaccgggc  tgtccctgcc  cagcaccatg
    30  16301  gtcttcgact  acccgacacc  gctggaactc  gcccagtacc  tgcgggcgga
        16351  gatcctcggc  gcggtgctgg  aagtcgccgg  cccggtcgcc  accggcggcg
        16401  ccgacgacga  gccgatcgcc  atcatcggca  tggcctgccg  cttccccggc
        16451  ggcgtcagct  ccccggaaca  gctgtgggac  ctggtcgcct  ccggcaccga
        16501  cgcgatcagc  gagttccccg  tcaaccgcgg  ctggcagacc  gggcacctct
    35  16551  tcgacccgga  ccccgaccgg  cccggcacca  cctactccac  ccagggcggc
        16601  ttcctccacg  aggccgacga  gttcgacccc  accttcttcg  gcatctcgcc
        16651  ccgcgaggcg  ctggtcatgg  acccgcagca  gcggctcctg  ctggagacca
```

```
      16701  cctgggagtc cttcgagcgc gccgggatcc gcccggaaac cctccgatcc
      16751  accctgaccg gcaccttcgt cggctccagc taccaggagt acggcctggg
      16801  cgcgggcgac ggcaccgagg gccacatggt caccggcagc agccccagtg
      16851  tgctctccgg ccgactgtcg tacgtcttcg gtctggaagg cccggcggtc
    5 16901  accgtcgaca ccgcctgctc gtcctcgctc gtggcgctgc acctggcctg
      16951  ccagtcgctg cgcaacggcg agagcaacct ggccgtcgcc ggcggcgcca
      17001  cgatcatgac gacgcccaac ccgttcatcg cgttcagccg gcagcgcgcc
      17051  ctcgccaagg acggccgctg caaggcgttc tccgacgacg cggacggcat
      17101  gacgctcgcc gagggcgtcg gcgtcgtcct cgtcgagcgg ctctccgacg
   10 17151  cgcagcgcaa cggccacccg gtcctggccg tcctccgcgg ctccgccatc
      17201  aaccaggacg gcgcctccaa cggcctgacc gcgcccaacg gcccgtccca
      17251  gcagagggtc atccgccagg ccctcgccaa cgcccgcctc gcgcccgggg
      17301  acatcgacgc cctggaggcg cacggcaccg gcacaccgct cggcgacccc
      17351  atcgaggccc aggcactgtt cgccacctac ggccgcgacc gcgaccccga
   15 17401  gagcgcgctg ctgctcggct cggtgaagtc caacatcggc cacacccagt
      17451  ccgccgcggg catcgccagc gtcatcaaga tggtcatggc gctgcgccac
      17501  tccgaactgc cgccgaccct gcacgccgac gcgccgtcct cgcacgtgga
      17551  ctggtcggcc gggacggtcc ggctgctgac ccaggcgcgc gcctggccgg
      17601  agaccggtcg cccgcgccgg gccgcggtgt cctcgttcgg catcagcggt
   20 17651  accaacgccc atgtcctgct ggagcaggcg cccgtcgcgg acaccccggc
      17701  cgaggagcgg cccgccgtgg cgccggtccc gatcgccgcc ggcgtcgtcc
      17751  cgtgggtggt caccgcccgc agcgccgccg ccctgcgcgg ccaggccgag
      17801  cgcctcctcg cgcacgccga aaccgtcgga accgccctgc cggccgccgg
      17851  accgctcgac atcggcctgt cgctggtctc cgcgcgcgcc cgtttcgagc
   25 17901  accgtgccgt cgtcgtcccg cccgcgggca ccgaccgct ggccgcgctg
      17951  cgcgccgtcg cgacggacgg gccctcgccc gtggtcgccc gtggcgtggc
      18001  ggacgtcgag ggtcggacgg tgttcgtgtt ccccggtcag ggttcgcagt
      18051  gggtggggat ggggtcccaa ctccttgatg agtcggcggt gttcgcggag
      18101  cggattgccg agtgtgcggc ggcactcgcc gagttcaccg actggtcgct
   30 18151  ggtcgatgtg ctgcggggtg tggtgggtgc gccgtcgttg gagcgggtcg
      18201  atgtggtgca gccggcgtcg ttcgcggtga tggtgtcgtt ggctgcgttg
      18251  tggcgttccc gtggtgtgtt gccggatgcg gtggtggggc attcgcaggg
      18301  tgagatcgct gctgcggtgg tgtcgggtgc gctgtcgttg cgggacgggg
      18351  cgcgggtggt ggcgctgcgg agtcaggcca ttggtcgtgc gttggcgggg
   35 18401  cggggaggga tgatgtccgt cgcgctgtcg gtggacgtgc tcgaaccgcg
      18451  gttggtcgag ttcgaggggc gggtgtcggt ggccgccgtc aacggcccgc
      18501  gctccgtcgt ggtcgccggc gagcccgagg cgctggacgc gctgcacgcc
```

```
18551   cggctgaccg  ccgacgacat  ccgggcccgc  cggatcgcgg  tggactacgc
18601  .ctcgcactcg  caccaggtcg  aggacctgca  cgaggaactg  ctggaggtgc
18651   tggcggagct  ggcgccgcgc  acgtcggagg  tgccgttctt  ctcgaccgtg
18701   accggcgact  ggctggacac  cgcgcggatg  gacgccggct  actggttccg
18751   caacctgcgc  ggacgggtgc  ggttcgcgga  cgcggtggcg  gacctgctgg
18801   cggcggagta  ccgcgcattc  gtcgaggtca  gctcgcaccc  ggtgctgtcg
18851   atggcggtgc  aggaggcgat  cgacgaggcc  ggcgtgccgg  ccgtcgccgc
18901   cggcaccctg  cgccgcgacc  agggcggcac  cgaccgcttc  ctgctgtcgg
18951   ccgccgaggt  cttcgtgcgc  ggtgtggacg  tggactgggc  ggggctgttc
19001   gaggggaccg  gtgcgtcccg  gatcgacctg  cccacctacg  ccttccaaca
19051   cgaacacctg  tgggccgtcc  cgcccgcccc  ggaggccgtc  gccgccgccg
19101   acccggacga  cgcggccttc  tggaccgcgg  tcgaggacgg  tgacgtctcc
19151   gcgctgaccg  ccgcgctcgg  caccgacgag  gactccgtcg  ccgccgtgct
19201   gcccgccctg  acctcctggc  gccgggcccg  ccgcgaccgc  tccaccgtgg
19251   acgcctggcg  ctaccgcgtc  gcctggaaac  ccctcggcgg  caccctgccg
19301   cacccgtccc  tgaccggcac  ctggctgctg  gtcaccgccg  acggcatcga
19351   cgacaccgat  gtggcagggg  cgttggagac  ctacggcgcc  gaggtgcgcc
19401   ggctggtcct  ggacgaggag  tgcgtcgacc  gcgccgtcct  gcgggagcgg
19451   ctggccggcg  cggaggacgt  gaccggcatc  gtctccgtcc  tcgccgccgc
19501   cgagcggacc  gacgcggtac  cgggcacctc  cctggtgctc  ggcaccgccc
19551   tgaccgtggc  actgatccag  gccctcggcg  acgccgaaat  cgacgctccc
19601   gtatgggcgt  tgacccgcgg  cgcggtctcc  accggccggg  ccgacgagct
19651   gaccgcgccc  gtccaggcac  aggtcaccgg  catcggctgg  accgcggcgc
19701   tggagcaccc  gcagcgctgg  ggcggcaccc  tcgacctgcc  cgccgccctc
19751   gacgcccggg  ccgcccagcg  gctcgccgcc  gtgctgtccg  gcgccctcgg
19801   cagcgacgac  cagctggcca  tccggccctc  cggggtcttc  acccgccgca
19851   tcgtgcgggc  cgaggccacc  gccgggcggc  cgccggcac  ctggacgccg
19901  .cgcggcacca  cactggtcac  cggcggctcc  ggcaccctcg  ccccgcacct
19951   cgcccgctgg  ctggcccaac  gcggcgccga  gcacctggtc  ctgatcagcc
20001   ggcgcggcac  ggccgccccg  ggcgccgccg  aactcgtcgc  ggaactggcc
20051   gagtcgggca  ccgaggcgac  cgtcgccgcc  tgcgacatca  ccgaccgcga
20101   cgcggtcgcc  gcgctgctgg  ccgacctcaa  ggccgacggg  cgcaccgtcc
20151   gcaccgtcgt  gcacaccgcc  gccaccatcg  agctgcacac  cctggacgcc
20201   accaccctgg  cggacttcga  ccgggtgctg  cacgccaagg  tcaccggcgc
20251   ccaggtcctc  gccgaactgc  tcgacgacga  agagctggac  gacttcgtcc
20301   tgtactcctc  caccgccggc  atgtggggca  gcggcgccca  cgccgcctac
20351   gtcgccggca  acgcctacct  cgccgcgctc  gccgagcacc  gccgggccaa
```

```
20401  cggactgccc  gccctgtcgc  tgtcctgggg  catctgggcc  gacgacctca
20451  aactgggccg  ggtcgatccc  cagatgatcc  ggcgcagcgg  cctggagttc
20501  atggacccgc  agctggccct  gagcggcctg  cagcgggcgc  tggacgacaa
20551  cgagaacgtg  ctcgcggtcg  ccgacgtgga  ctgggagacc  taccaccccg
20601  tctacacctc  cggccgaccc  accccgctct  tcgacgaggt  gccggaggtc
20651  cgccggctca  ccgcggccgc  cgagcagagc  gccgggaccg  tcgccgaggg
20701  cgagttcgcc  gccgcgctgc  gcgccctgtc  cgacgccgag  cagcagcgca
20751  ccctgctgga  gaccgtccgc  accgaggcgg  cgtccgtcct  cgggctgtcc
20801  tccgccgagg  acctcaccga  ccagcgggcc  ttccgcgacg  tcggcttcga
20851  ctcgctgacc  gccgtcggcc  tgcgcaaccg  gctcgcctcc  gtcaccggcc
20901  tgacgctgcc  ctcgacgatg  gtcttcgact  accccaaccc  ggccgcgctc
20951  gccgcctatc  tgcacggcga  gctggccggc  gcccggtccg  ccgccgccgg
21001  cgccgccgcc  gtcccgaccg  gcgccccga  cgccgacgac  ccgatcgcga
21051  tcgtcggcat  gagctgccgc  taccccggcg  gggtcggctc  cgccgaggac
21101  ctgtggcgga  tcgccctgga  cgaggtcgac  gcgatctccg  gcttccccgc
21151  cgaccgcggc  tgggacgccg  agggcctcta  cgacccggac  cccgaccggc
21201  ccggccgcac  ctactccgtc  cagggcggat  tcctgcgcga  cgtcgccgag
21251  ttcgacccgg  gcttcttcgg  gatctcgccg  cgcgaggcgc  tgtcgatgga
21301  cccgcagcag  cggctcctgc  tggagaccgc  ctgggaggcg  ttcgagcacg
21351  ccggcatcga  cccggtcggc  cagcgcggca  gccgcaccgg  caccttcgtc
21401  ggcgccagct  accaggacta  cgcctccggc  gtgcccaaca  gcgagggctc
21451  cgaaggccac  atgatcaccg  gcacgctctc  cagtgtgctg  tccggccggg
21501  tgtcctacct  cttcggcttc  gaggcccccg  ccgtcacgct  cgacaccgcc
21551  tgctcctcct  ccctggtcgc  catgcacctg  gcctgccagt  ccctgcgcaa
21601  cggggagagc  tcgctggccc  tggccggcgg  cgtcagcatc  atgtccaccc
21651  cgatgtcgtt  cgtcggcttc  agccggcagc  gcgccctcgc  cgaggacggc
21701  cgctgcaagg  cgtacgcgga  cggcgccgac  gggatgaccc  tcgccgaggg
21751  cgtcggcctg  gtgctgctgg  agcggctgtc  cgacgcccgc  gccaacgggc
21801  accaggtgct  cgccgtgatc  cgcggctccg  ccgtcaacca  ggacggcgcc
21851  tccaacggcc  tgaccgcacc  caacggcccg  tcccagcagc  gcgtcatccg
21901  ccaggcgctg  gccaactccg  cggtggcgcc  cggcgacatc  gacgtcctgg
21951  agggccacgg  caccggtacc  gccctcggcg  accccatcga  ggcgcaggcc
22001  ctgctcgcca  cctacggcca  ggaccgcgcc  cccgaacggc  cgctgctgct
22051  cggctcggtg  aagtccaaca  tcggccacac  ccagatggca  tccggcgtcg
22101  ccagcgtcat  caagctcgtc  cgcgccctcc  aggaaggcgt  ggtgcccaag
22151  tccctgcaca  tcgaccggcc  ctccacccac  gtcgactggt  cctcgggcgc
22201  catcgggctg  ctcaccgaac  gcaccccgtg  gcccgagacc  ggccggccgc
```

```
20401  cggactgccc  gccctgtcgc  tgtcctgggg  catctgggcc  gacgacctca
20451  aactgggccg  ggtcgatccc  cagatgatcc  ggcgcagcgg  cctggagttc
20501  atggacccgc  agctggccct  gagcggcctg  cagcgggcgc  tggacgacaa
20551  cgagaacgtg  ctcgcggtcg  ccgacgtgga  ctgggagacc  taccaccccg
20601  tctacacctc  cggccgaccc  accccgctct  tcgacgaggt  gccggaggtc
20651  cgccggctca  ccgcggccgc  cgagcagagc  gccgggaccg  tcgccgaggg
20701  cgagttcgcc  gccgcgctgc  gcgccctgtc  cgacgccgag  cagcagcgca
20751  ccctgctgga  gaccgtccgc  accgaggcgg  cgtccgtcct  cgggctgtcc
20801  tccgccgagg  acctcaccga  ccagcgggcc  ttccgcgacg  tcggcttcga
20851  ctcgctgacc  gccgtcggcc  tgcgcaaccg  gctcgcctcc  gtcaccggcc
20901  tgacgctgcc  ctcgacgatg  gtcttcgact  accccaaccc  ggccgcgctc
20951  gccgcctatc  tgcacggcga  gctggccggc  gcccggtccg  ccgccgccgg
21001  cgccgccgcc  gtcccgaccg  gcgccccga  cgccgacgac  ccgatcgcga
21051  tcgtcggcat  gagctgccgc  taccccggcg  gggtcggctc  cgccgaggac
21101  ctgtggcgga  tcgccctgga  cgaggtcgac  gcgatctccg  gcttccccgc
21151  cgaccgcggc  tgggacgccg  agggcctcta  cgacccggac  cccgaccggc
21201  ccggccgcac  ctactccgtc  cagggcggat  tcctgcgcga  cgtcgccgag
21251  ttcgacccgg  gcttcttcgg  gatctcgccg  cgcgaggcgc  tgtcgatgga
21301  cccgcagcag  cggctcctgc  tggagaccgc  ctgggaggcg  ttcgagcacg
21351  ccggcatcga  cccggtcggc  cagcgcggca  gccgcaccgg  caccttcgtc
21401  ggcgccagct  accaggacta  cgcctccggc  gtgcccaaca  gcgagggctc
21451  cgaaggccac  atgatcaccg  gcacgctctc  cagtgtgctg  tccggccggg
21501  tgtcctacct  cttcggcttc  gagggccccg  ccgtcacgct  cgacaccgcc
21551  tgctcctcct  ccctggtcgc  catgcacctg  gcctgccagt  ccctgcgcaa
21601  cggggagagc  tcgctggccc  tggccggcgg  cgtcagcatc  atgtccaccc
21651  cgatgtcgtt  cgtcggcttc  agccggcagc  gcgccctcgc  cgaggacggc
21701  cgctgcaagg  cgtacgcgga  cggcgccgac  gggatgaccc  tcgccgaggg
21751  cgtcggcctg  gtgctgctgg  agcggctgtc  cgacgcccgc  gccaacgggc
21801  accaggtgct  cgccgtgatc  cgcggctccg  ccgtcaacca  ggacggcgcc
21851  tccaacggcc  tgaccgcacc  caacggcccg  tcccagcagc  gcgtcatccg
21901  ccaggcgctg  gccaactccg  cggtggcgcc  cggcgacatc  gacgtcctgg
21951  agggccacgg  caccggtacc  gccctcggcg  accccatcga  ggcgcaggcc
22001  ctgctcgcca  cctacggcca  ggaccgcgcc  cccgaacggc  cgctgctgct
22051  cggctcggtg  aagtccaaca  tcggccacac  ccagatggca  tccggcgtcg
22101  ccagcgtcat  caagctcgtc  cgcgccctcc  aggaaggcgt  ggtgcccaag
22151  tccctgcaca  tcgaccggcc  ctccacccac  gtcgactggt  cctcgggcgc
22201  catcgggctg  ctcaccgaac  gcaccccgtg  gcccgagacc  ggccggccgc
```

```
       22251  gccgcgccgc cgtctcctcc ttcggcatca gcggcaccaa cgtccacacc
       22301  atcctcgaac aggcccccgc ggacgaggcg cccacgcccg ccgacccgcc
       22351  gcgggacggc ctggtgccgg tcctgctctc cggccgcggc gaggccgcgc
       22401  tgcgcgccca ggccgcccgc ctgctcgcct tcgtcgagga gcggcccgag
   5   22451  gcccacctca ccgacctcgc ccactccctc gccacctcgc gcgccgcgct
       22501  ggaacgccgc gccgccgtca tcgccgccga ccgcgacacc ctgacccgcg
       22551  gcctgcgcgc cctgtccgac ggccggcccg accccggcct ggtccagggc
       22601  accgcgggac gcggccggac cgccttcctg ttcaccggac agggcagcca
       22651  gcgccccggc atgggccgcg aactccacga ccgctacccg gtgttcgccg
   10  22701  acgcgctgga cgaggtgctg gcccggctcg acgacggacc ggaccggccg
       22751  ctgcgcgagg tgctgttcgc cgcgcccgac tccgccgagg ccgcgctcct
       22801  ggaccggacc ggctacgccc agcccgcgct gttcgccgtc gaggtcgcgc
       22851  tgttccgcct gctgacgtcc tggggcctga ccccggacta cctggccggc
       22901  cactccgtcg gcgaactcgc cgccgcgcac gtcgccggcg tgctgtcgct
   15  22951  ggacgacgcc tgcactctgg tcgccgcccg cggccggctc atgcaggcgc
       23001  tgcccgaggg cggcgcgatg gtcgccctgg aggccgcgga ggacgaggtc
       23051  ctgccgctcc tggaggggct caccgaccgg gtgtccgtcg ccgccgtcaa
       23101  cgggccgcgg tccgtggtcg tcgccggcgt cgaggaggac gtgctcctcc
       23151  tcgccgacct cttcgccgcc gacgggcgcc ggaccaagcg gctgcgggtg
   20  23201  agccatgcct tccactcgcc gctgatggac gccatgctcg acgacttcgc
       23251  cgccgtcgcc cgcgggctga cctaccaccc gccgacgatc ccgttcgtgt
       23301  cgaacgtcag cggcggcctg gccaccgccg aacaggtccg cacgcccgac
       23351  tactgggtcg ggcacgtccg cgccgcggtc cgcttcgccg acggcatcga
       23401  ctggctcgcc acccagggcg acgtccacac cttcctggag ctcggcccgg
   25  23451  acggcgtgct cagcgccatg gcccgggaga gcctcaccga cccgtcccgc
       23501  acggcactgc tgccgaccct gcgcggcgac cggcccgagg aacctgccct
       23551  ggtcaccgcg gtcgccgcgg cccacgcgca cggcgcccgc gtcgactgga
       23601  gcgggtactt cgccgaccac ggcgcgcgcc ggaccacgct gccgacctac
       23651  gcgttccaac gcgagcggta ctggcccgac accaccgccg ccacgagcgc
   30  23701  ccacgcgccc ggatccgccc tcgacgccga gttctgggcc gccgtcgagc
       23751  gggacgacgt cgccgccctc gccgcctccc tggacctgga cgacgccacc
       23801  gtcaccgcga tggtccccgc gctcaccgcc tggcgccggc gccgcggcga
       23851  gcagaccgaa ctggactcct ggcgctaccg cgtcacctgg aagccgcgcg
       23901  gcggcgccac cgcacccgcc gccctcaccg gccgctggct cgtgctcgtc
   35  23951  ccgcacgacc accaggaccg tcaggacgac gcgaccgcgg cctgggcagc
       24001  cgacgtcgag accgccctgg gcaccaccac cgtccggctg acggtcacca
       24051  ccaccgaccg cgccgcgctg gccgcccgga tcaccgaagc cgccggcgac
```

```
24101  cagggcccgt tcagcggtgt gctgtccctg ctgccgctcg ccaccggcga
24151  cgccggccac cccggtgcgc ccgccgccct caccctcacc accaccgccg
24201  tccaggccct cggcgacgcc ggcatcgacg cgccgctgtg gaacgtcacc
24251  cgcggggccg tggccgtcgg ccgcgccgaa caggtcaccg cgcccgaaca
24301  ggccgccgtc tggggcctgg gccgcgccgt cgccctggaa ctgccggccc
24351  ggttcggcgg caccctcgac ctgcccgcca ccctggacgg ccaggccgcc
24401  cgccggttgc gcgcggtgct cgcggctacc gacggcgagg acgcggtggc
24451  cctgcggccc tccggcgtct tcctccgccg cctggcccac gccccggccg
24501  gccccgacac cgcccgcacc gccttcgacc cggcggccgg caccgtcctg
24551  atcaccggcg gcaccggcgg catcggcggc cacgtcgccc gccgcctggc
24601  ccgcgacggc gccacccacc tgctgctcac cagccgccgc ggcccggccg
24651  cccccggcgc cgacgcgctc cgcgccgaac tggaggaact gggcgcccgg
24701  gtcaccctcg ccgcctgcga cgccgccgac cgcgacgcgc tggccgcgct
24751  cctcgccgaa ctgcccgacg acgccccgct gtgcgcggtg ttccacaccg
24801  ccggcgtcgt cgaggaccac gtcgtggacg cgctcacacc ggagaacttc
24851  gccgcggtgc tgcgcgccaa gaccgtcgcc gcccaccacc tgcacgagct
24901  gaccgccgac ctggacctcg ccgctttcgt gctgttctcc tccacggccg
24951  gcgtcctcgg cgccgccgga cagggcaact acgccgcggc caacgcccac
25001  ctggacgccc tcgccgaaca ccgccgctcc cacggcctga ccgcgctgtc
25051  cgtcgcctgg ggcccgtggg ccggctccgg catggtcgcc gacgccgccg
25101  aactcaccga ccgggtacgg cgcggcggct tcgaaccgct cgcccccgaa
25151  ccggccgtgc gcgccctgct gcgcgccatc gagaacgacg acaccaccgt
25201  cgcgctcgcc gacatcgact gggagcgctt ccagcgcgcc ttcgccgcgg
25251  tccgcccgct gccgttcgtc gccgacctcc ccgagaccgg ccgggccacc
25301  cccgcgaccg ccaccggcgc cgccaccggc ctgcggcagc aactcgccga
25351  actgccggag cacgagcgcc cggcagcggt cctggacctg ctgcgtaccc
25401  aggtcgccgc cgtcctcggc cacgccgacc cgcgcaccgt cgaggacgac
25451  cacgccttcc gcgacctggg cttcgactcg ctgaccatcc tggaactgcg
25501  caacgccctc aacgccgcca ccggcctgag cctgcccgcc accctggtct
25551  acgacctgcc caccccgcgc gagatggcg acttcctgct cgccgaactc
25601  ctcggcaccc tgcccaccga caccgccgcg accgtcgcca gcacggcctc
25651  ccccaagctc tcagcttcgt tcgagcaggg cggtaccccc ttcgacgacc
25701  cgatcgccgt catcggcatc ggctgccgct tccccggcgg cgtcaccacc
25751  ccggaggagc tctggcagct cctcgacgag ggccgcgacg gcatcagccg
25801  cttccccgac gaccgcggct gggacctcgc cgcgctgggc gccggcgcct
25851  ccgacaccct ggagggcggc ttcctgaccg gcgtcgccga cttcgacgcc
25901  cggttcttcg gcatctcgcc ccgcgaggcg ctggccatgg accccccagca
```

```
     25951  gcggctgctg ctggagacca cctgggaggc gctggagcgg gccggcatcg
     26001  acccgaccac gctgcgcggc tccaccaccg gcgtcttcgt cggcaccaac
     26051  ggccaggact acccgacgct gttgcgccgc tccgcctcgg acgtggccgg
     26101  ctacgtcgcc accggcaaca ccgccagcgt gatgtccggc cgcctgtcct
  5  26151  acgcgctcgg cctcgaaggc ccggccgtca ccatcgacac cgcctgctcc
     26201  tcctcgctcg tcgccctgca ctgggccggc cgggcgctgc gcgccggcga
     26251  gtgcgacctc gtggtggccg gcggcgtctc ggtcatggcc agcccggact
     26301  ccttcgtcga gttctccacc cagggcggcc tggcacccga cgggcgctgc
     26351  aaggcgttct ccgacgccgc cgacggcacc gcctggtccg aaggcgtcgg
 10  26401  catcctcgtc ctggaacgcc tctccgccgc ccgccgcaac ggccaccagg
     26451  tcctcggcct gatccgcggc accgccgtca accaggacgg cgcgtccaac
     26501  ggcctgaccg cgcccaacgg cctctcccag cagcgcgtca tcgcccaggc
     26551  actcgccgac gcccgcctgc gccccgccga catcgacgcg atcgaggcgc
     26601  acggcaccgg caccaccctc ggcgacccga tcgaggcccg cgccctgatc
 15  26651  accgcctacg gccgggaccg ggacgccgaa cggccgctgc tgctgggcac
     26701  cgtcaagtcc aacatcgggc acacccaggc cgccgccggt gccgccggcg
     26751  tcatcaagat gctgatggcg atgcgccacg gcaccctgcc caggacgctg
     26801  cacgtgggca ccccgtccag ccacgtcgac tggagcggcg gcaccgtcgc
     26851  gctcctcgac gacgcgcggc cctggccacg gaccgggcag ccgcggcgcg
 20  26901  ccggcgtctc cgccttcggc gtcagcggca ccaacgccca cgtcgtcgtc
     26951  gagcaggccc cggaaaccga agcccccgcc gccccggccg ccgagccggc
     27001  gccggaggcc acgcccaccg tcgtcccctg ggtcgtctcc ggacgcagcc
     27051  gggaagcgct ccaggcgcag ctggaccggc tcaccgcgca caccgccgcc
     27101  caccccgcgc gctcggcggc ggacgtcggc cgctcgctgg ccaccgaccg
 25  27151  cacgctcttc ccgcaccgcg ccgtgctgct cgccggcccg gacggggtgc
     27201  gcgaggccgc ccgcgccgcc gcgccccgca cccccggccg caccgcgttc
     27251  ctgttctccg gacagggcgc ccagcacgcc ctgatgggcc acgacctgta
     27301  ccagcgcttc ccggtctacg ccgacgcact ggacaccgtc ctcgcccagt
     27351  tcgacaccgt gctggacgtc ccgctgcgcg ccgcgctgtt cgccgcgccg
 30  27401  ggcacccccg aggccgcgct cctggaccag accggcttca cccagcccgc
     27451  gctgttcgcc gtcgaggtcg cactgttccg gctcgccgag tcctggcggc
     27501  tgacgccgga cttcgtcgcc ggccactcca tcggcgagat cgccgccgcg
     27551  cacgtcgccg gggtgttctc cctggaggac gcctgcacgc tggtcgccgc
     27601  ccgcgcctcc ctcatgcagc aactgccgcg ggacggtgcg atggtggccc
 35  27651  tggaagccac cgaggacgag gtcgcgccgc tgctcaccga cggcgtcgca
     27701  ctcgccgcgg tcaacggccc ccgctcggtg gtcgtcgcgg gcgccgagga
     27751  cgccgtccgc gcggtcgccg accggctcgc cgccgacggc cgccgcaccc
```

```
       27801  gccggctgac ggtcagccac gccttccact cgccgctgat ggacccgatg
       27851  ctcaccgact tcgcccgggt cgcggagggc ctgacctacc acgagccgcg
       27901  catcccctc gtctccaccc tcctcggcgc cccggccggc gccgaactgc
       27951  gcacccccga ctactgggtg cggcacgtcc gcgagaccgt gcggttcgcc
   5   28001  gacggcgtgc gcgccctgca cgacgccggc gccggcacct cgtggagat
       28051  cggcccggac ggcgtgctca ccgccctgac ccagcagacc ctcgacaccg
       28101  tcgaggccgg cgcgcccgcc gtcgtcgtgc cgctccagcg ccgcgaccgc
       28151  gccggcgacc tcgccctcct ggagggcctg gccaccctgc acacccacgg
       28201  caccggcccg tcctggcccg cctacttcga ggccaccggc ggccaccgga
  10   28251  ccgatctgcc cacctacgcc ttccagcggg agcggtactg gcccgaactc
       28301  ggcgcacccg tcgccaccgc cccgcaggac ccggcggcct ggcgctacca
       28351  cgagacctgg gccccgctgc cggcacccga ggcggccgcg ccggccggcc
       28401  gcgccctggt cctcgtcccg gccgggaacc gcgacaccgc gtggatgacg
       28451  gccgtcgccg acgcgctcgg cgccgacacc gtcacggccg aacccgacgc
  15   28501  cctggccgag cagttgacgg ccgccggcga cacaccctgg cgcgtcgtgg
       28551  tgtcgctgct cgccgccgcg tccgaggggc tgcccgcgga cggcgcctgg
       28601  cccgccgccc tcctcgccac cctggacgag gccggcgtgc acgcgcccct
       28651  gtggtgcgtc acccgcggcg ccgtcgcggt cgcggggggag gccccgaccg
       28701  ccgtcggcca ggccgccctg tggggcctgg ccgggtcgc cgcgctggac
  20   28751  cacccggacc gcttcggcgg cctggccgac ctgcccgccg acaccgacgc
       28801  gcacgccgcc gggctgctcg ccgcgcacct ggccgcgccg ggcaccgagg
       28851  ccgagatcgc ggtccgcgcc accggcgtcc acgcccgtcg cctggtccgt
       28901  acgccggccg ccgccgacgg tgccacctgg ctgccgaccg gcaccgtcct
       28951  ggttgtcggc ggcaccggcg gcaccggcac catgggcggc cgggccgccc
  25   29001  gctggctggt ccgcgagggc gcccgccacc tcgtcctgac cgccccccgac
       29051  ggcaccacga ccgccgcgga caccgaggcc ctgacggccg aactggccgc
       29101  gctcggcgcc cggatcaccg tcgtggacca cgaccccacc gccccggacg
       29151  gcttcgccgc gctcctcgac ggactgcccg acgacacccc gctcaccgcg
       29201  gtcgtgtacg cgccggaggc cgacgccgcc cccggcaccg cggccgagct
  30   29251  gtccgccgca ctcgcccccg tcaccgccct aggcgccgcc ctcaccggcc
       29301  ggccgctgga cgccttcgtc ctcttcggct ccatcgccgg gctctggggc
       29351  gtgcgcggcc gggccgccga ggccgcgtcc ggcgcctacc tcgacgcctt
       29401  cgcccgcgcc tgccgcgacc gcggcacccc ggcactggcc gtcgcctggg
       29451  gcgcctgggc cgacctggtc ggcccgtccc tcgccgcgca cctgcggatg
  35   29501  aacggcctgc cggtgatgga cgcggacacc gcactgaccg ccctcagccg
       29551  ggccgtcgcc gacgggtccg ccgccgaggc ggtcgccgac gtccgctggg
       29601  agaccttcgc gccccctccac cacgaggccc gccgcaccgc cctgttcgac
```

```
       29651  gccctgcccg aggcccgcgg cgcgctcgcg gaggccgccc gggaccgcgc
       29701  cgaccggaag accgccgccg gcgactacgg ccggtggctc gccgagcagc
       29751  ccgccgcgga ccacgacgcc atcctgctgg cactggtcac cgagaaggcc
       29801  gcgaccgtcc tcggccacgc cgaccacgac ctgctcgaac ccgacctgcc
    5  29851  cttccgcgac ctgggcttcg actcgctgac cgcggtcgac ctgcgcaacc
       29901  agctcaccgc ggaaaccggc ctcaccctgc ccgccaccct cgtcttcgac
       29951  caccccaacc cggccgccct cgccgcccac ctgcgcgccc aactcctcgg
       30001  cgaggcgagc gactccgccg caccggtggc cgcccccgtc gccctcggtg
       30051  ccgacgacga cgcgatcgtc atcgtcggca tggcctgccg ctaccccggc
   10  30101  ggggtcacct cgcccgagga cctgtggcag ctggtcggcg acgaggtcga
       30151  cgcggtcggc gacttcccga ccgaccgcgg ctgggacctg gccgcgctcg
       30201  ccggcgacgg accgggccgc agtgccaccg cccagggcgg attcctctac
       30251  gacgccaccg acttcgaccc cggcctgttc ggcatctcgc cgcgcgaggc
       30301  cctggtgatg gacccgcagc agcggatcct gctcgaaacg tcctgggagg
   15  30351  ccctggagcg ggccggcatc gacccggcga cgctgcgcgg cagcggcacc
       30401  accggcgtct tcgtcggcgg cggctccggc gactaccggc cgccggagga
       30451  ggccgggcag tggcagaccg cccagtccgc cagcctgctc tccggtcgcc
       30501  tcgcctacac cttcggcatc cagggcccca ccgtgtcggt cgacaccgcc
       30551  tgctcctcgt cgctggtcgc gctgcacctg gccgcgcagg ccctgcgcgc
   20  30601  cggcgaatgc tcgatcgcgc tggccggcgg cgtcaccgtg atggccaccc
       30651  cggtgggctt cgtcgagttc agcgcccagg gcgccctgtc gccggacggc
       30701  cgctgccgcg ccttctccga cgacgccaac ggcaccggct ggtccgaagg
       30751  cgtgggcatg ctcgtcgtcg aacggctctc cgacgcccgc cgcaacggcc
       30801  accgcgtcct cgccgtgctc cgcggctccg ccatcaacca ggacggcgcg
   25  30851  tccaacggcc tgaccgcccc cagcggcccc gcccagcagc gcgtcatccg
       30901  ccaggccctc gccaacgccc gactgcgccc cgccgacatc gacgccgtcg
       30951  aggcccacgg caccggcacc aggctcggcg accccatcga ggcccaggcc
       31001  ctgctcgcca cctacggcca ggaccgcgag cggcccgtgc tgctcggctc
       31051  gctcaagtcc aacatcggcc acacccaggc cgcctccggc gtcggcggcg
   30  31101  tcatcaagat ggtcctcgcc atgcagcacg gcgaactgcc gcgctccctg
       31151  tacgccgaga cccctcgtc gcacgtggac tggaccgccg gccgcgccca
       31201  cctgctcacc gccaggaccc cgtggcccga ctccggtcgg ccgcgccgcg
       31251  ccgccgtctc ctccttcggc gccagcggca ccaacgccca cgccatcctg
       31301  gagcagccgc cgcgcgagga actccccgcg cgccccgcgg acgacggcgc
   35  31351  cccgctgccg ttcttgctct ccggccgctc gcagaacgcc ctgcgcgccc
       31401  aggcccgccg actcctggcc cgcctcaccg cccaccccga cacccgggcc
       31451  gccgacctgg cgtactccct ggcgaccacc cgggccgcct tcgagcaccg
```

SUBSTITUTE SHEET (RULE 26)

```
        31501  ggccgcgatc accgccaccg accacgacgg cctccgcacc ggcctgaccg
        31551  ccgtcgccga gggcaccacc gccccgcaca ccgccgaaca ccacctccag
        31601  ggcaccggaa agcgcgccgt gctcttctcc ggccagggct cccagcgcct
        31651  gggcatgggc cgcgaactgc acgagcgcca cccggtgttc gccgaggcgt
    5   31701  tcgactccgt actggcccgc ctcgacgacc ggctcgacac ccccctgcgg
        31751  gacgtcgtct ggggcaccga cgaggaggcg ctgcacgcca ccgggaacac
        31801  ccagcccgcc ctgttcgccg tcgaagtcgc gctctaccgc ctgatcgaat
        31851  cctggggcgt gcggcccgac ttcgtggccg gccactccgt cggcgagctc
        31901  gccgcggccc acgtcgccgg ggtgctctcc ctggacgacg cctgccgcct
    10  31951  ggtcgccgcc cgcgccgccc tcatgcagcg cctcccggcc ggcggcgcca
        32001  tgatcgccgt cgaggccacc gaggacgagg tcaccccgct cctcaccgac
        32051  ggcgtgtccc tcgccgcggt caacggaccg accgccgtgg tcctctccgg
        32101  cgcgggcgac gccgtgaccg ccctgggcca ggcgctggcc gaacggggcc
        32151  accgcaccac ccggctgcgg gtcagccacg ccttccactc gcacctcatg
    15  32201  gacccgatgc tggcggactt ccgcaccgtc gccgagggcc tggaatacca
        32251  cccgccgcgc atccccgtgg tctccaacct caccgggggac gtcgccgacg
        32301  cggccgacct gtgctccgcc gactactggg tgcgccacgt ccgcggcacc
        32351  gtacggttcg ccgacggcgt gcgcaccatg gccgaccgcg gcgtgcacct
        32401  cttcctcgaa ctcggcccgg acgccgtgct gtcggccatg gcccgccagt
    20  32451  gcgcaccgga cgccgtcgtc gtcccggccc tgcgccgcaa ccgcgacgag
        32501  gacgagacgc tggtcggcgc cgtcgcgcga ctgcacgtcc acggcgcggg
        32551  tccgcgctgg gacgcgtact tcgccggccg cggcgcccag tggctggacc
        32601  ttccgacgta ccccttccag cgcggccgct tctggccgga gtcccttccg
        32651  ggcgccgcat cggccgcccc ggcagccgga cagccggccg agaccgacgc
    25  32701  ggccttctgg gacgccgtcg cacaggagga cttcaccgca ttggaatccg
        32751  tactcgacgt cgagagcgac gcactgtcca aggtgctgcc ggccctgatg
        32801  gactggcgca gccgccaggc cgacgagtcc caactggcag gctggcgcca
        32851  ccgcatcgtc tggaagcggc tcaccggcgc cgccctggca caccgcaagg
        32901  cgctcagcgg cacctggctc gcggtggtcc ccgagggctt cgccgacgac
    30  32951  ccctgggtga ccaccaccct ggacggcctc ggtacccacc tcgtgcatct
        33001  ggaggtcgcg gaggccgacc gggccgcgct ggccgacgcg atcgcggccc
        33051  gcaccgccga cggcacccgc ttcggcggcg taatctccct gctggccctg
        33101  cgcgaggagc tcaccggcgc ggtgcccgag gggaccgccc tgaccaccac
        33151  cctcctccag gccctcggcg acgccggcgt cgacgcaccg ctgtggtgcg
    35  33201  tcacccgcag cgccgtctcc gccggccgca ccgaccggcc gcaccgaccg
        33251  ctccaaggcg ccgtctgggg cctgggccgg gtcgcggccc ttgagtaccc
        33301  gcagcgctgg ggcggcctgg tggacctgcc ggaggagccc gacgagcggt
```

```
        33351  ccgcggccgg cctcgccgcc gtcctggccg gtctggacgg cgaggaccag
        33401  gtcgccgtgc gcggcaccgc ggtgctcgcc cgccgcctgg tgccggctcc
        33451  cggccgcaag ccgtcccggc cctggcaccc gtccggcacc gtcctggtca
        33501  ccggcggcac cggcgccctc ggcgcgcacg tcgcccgccg cctggccaag
    5   33551  gacggcgccc agcacctcgt cctgctcagc cgccgcggcc cggacgctcc
        33601  cggtgcggcg gaactgcgcg cggaactgga cgcgttgggc accgacgtca
        33651  cggtcgccgc ctgcgacgtc gccgaccgcg accagctgac ggccgtcctg
        33701  gacgcgctgc ccgccgaccg gccgctgacc ggtgtggtgc acaccgccgg
        33751  cgtcctcgac gacggcgtac tggaccggct caccccgag cggttccagg
    10  33801  aggtgttccg cgccaaggtc acctcggccc tgctgctgga cgagctgacc
        33851  cgcgaccgcg agctggccgc gttcgtcctc ttctcctccg cctccgccgc
        33901  ggtcggcaac ccgggccagg ccaactacgc cgctgccaac gccgtcctgg
        33951  acgcgctcgc cgaacagcgc cgggtgctcg gcctgcccgc cacctcggtc
        34001  tcctggggtg cctggggagg cggcggcatg gccgacgccg acggcgcgga
    15  34051  cgaggccgcc cggcgcgccg gcgtcggcgc catggacccg cacctcgccg
        34101  tggaagccct gctgcgcctg gtcgccgaga aggagccgac cgcggtggtc
        34151  gccgaggtgg ccctggaccg gttcgccggc gccttcggcg gcagccgacc
        34201  cagcgccctg ctgcgggagt tccccggcta ccgcgaggcg ctcgccgccc
        34251  aggcggagca ggccgcggac ggcggcgggc tggccgcccg actggccgcg
    20  34301  ctgccgcccg cccgccgcct ggacaccgtt gtggacctgg tgcgcacccg
        34351  cgccgcgcag gtgctcggct accccgacac cgaagcggtc gccgccgaac
        34401  ggtccttccg cgacctgggt gtcgactcgc tcggcgccgt cgagctgcgc
        34451  aaccaactga gcgcggccac cggcctgaac ctgccggcga cgctggtgtt
        34501  cgaccacccg accccctgg tcctggggga gcacatcctc ggcgggctct
    25  34551  tcccggacga gcccgccggg tccgacgacg agacggagat ccgggccctg
        34601  ctggcctccg tcccgctcga ccaactgcgg gagatcgggg tcctggagcc
        34651  cctgctccag ctcgccggac gcggcggccg ggccgcggac ggcgacgacg
        34701  gcgagtccgt cgactcgatg acagtggcag acctggtgcg ggccgcgctc
        34751  aacggccagt ccgacctgta gcgcgattga tggagcagac gatgaacgcg
    30  34801  cccgagaacc ccgagacccc cgagaacaac gtagtcgccg cactccgcgc
        34851  cgcggtcaag gagaccgacc ggctccggcg gcagaaccgg atgctggtcg
        34901  cggcggccaa ggaaccgatc gccgtggtcg gcatggcctg ccgcttcccc
        34951  ggcgccgtcg actccccgga agcgctgtgg gagatggtcg ccaccggcac
        35001  cgacgtgatc tccggattcc ccgacgaccg cggctgggac ctggaggcgc
    35  35051  tgcgcaacag cggcaccgac gcccgcgaca ccgacgtcag ccagcgcggc
        35101  ggattcctgg actgcatcgc cgacttcgac cccggcttct tcgggatctc
        35151  accgcgcgag gcggtcacca tggacccgca acagcggctc ctgctgacca
```

```
35201   ccgcctggga  ggccgtcgag  cgggccggca  tcgacgccac  cacgctgcgc
35251   gccacccgca  ccggcgcgtt  catcggcacc  aacggccagg  actacgccta
35301   cctgctcgtc  cgctccctgg  acgacgccac  cggcgacgtc  ggcaccggca
35351   tcgccgccag  cgccgcctcc  ggtcggctct  cctacaccct  cggcctcgaa
35401   ggccccgcgc  tcaccgtcga  caccgcctgc  tcctcgtcgc  tggtcgccct
35451   gcacctggcc  gtgcaggcgc  tgcgcaacgg  cgagtgcggc  atggcgctgg
35501   ccggcggcgt  caacgtgatg  gccacaccgg  gctcgctggt  cgagttcagc
35551   cgccagggcg  ggctggcccg  ggacggccgc  tgcaaggcgt  tcgcggacgc
35601   cgccgacggc  accggctggt  ccgagggcgc  cggcgtgctg  ctgctggaac
35651   ggctctccga  cgcccagcgc  aacggccacc  cggtgctcgc  cgtggtccgc
35701   ggctccgccg  tcaaccagga  cggcgcctcc  aacggcttca  ccgcccccaa
35751   cggcccctcc  cagcagcgcg  tcatccgcca  ggccctcgcc  aacgccggcc
35801   tggccaccgg  cgacatcgac  gcggtcgagg  cgcacgggac  cggcaccccg
35851   ctcggcgacc  ccatcgaggc  gcagagcatc  ctcgccacct  acggccagga
35901   ccgcgcccac  ccggtgctgc  tcggctcgat  caagtcgaac  atgggccaca
35951   cccaggccgc  gtccggcgtc  gccggcgtga  tcaagatgat  catggcgatg
36001   cggcacggcg  tcctcccgcg  gaccctgcac  gtcgaccggc  cctccaccca
36051   cgtcgactgg  accaccggca  gcgtcgaact  cctcaccgac  gcccacccgt
36101   ggcccgagac  tggcaggccg  cgccgcaccg  gcatctcctc  cttcggcgtc
36151   agcggcacca  acgcccacgt  catcgtcgaa  caggcccccg  acacccccgc
36201   cgaggcggct  gacgacactc  cgccccgcac  cccgcggacc  ctgccgtggc
36251   tgctctccgc  ccgcaccggc  gccgccctgc  gcgaccaggc  caccgcgctg
36301   ctcgaccacc  tcgaccgccc  cgacggcgac  cgcgggccca  ccgccctgga
36351   caccgcgttc  tccctcgcca  ccacccgcgc  cgccctggaa  caccggctcg
36401   ccgtcgtcac  cggcaccgac  ggcaccgccg  gacgggacgc  cctgaccgcc
36451   tggctggcgc  acggcaccgc  ccccgacgcc  cacgaaggac  acgccgccgg
36501   acgcacccgc  tgcgcggccc  tcttctccgg  ccagggcgcc  cagcgcctgg
36551   gcatgggccg  cgaactccac  gcccgtttcc  cggtgttcgc  acgggccctc
36601   gacaccgccg  tggacctgct  cgacgccgaa  ctgggcggca  ccctgcggga
36651   ggtgatctgg  ggcaccgacg  acgcgccgct  caacgagacc  ggcttcaccc
36701   agcccgccct  gttcgccgtc  gaggtcgccc  tctaccgcct  gatcgaatcc
36751   tggggcgtcg  ccccggactt  cgtcgccggc  cactccatcg  gcgagatcgc
36801   cgccgcgcac  gtcgccgggg  tgttctccct  ggaggacgcc  tgcacgctgg
36851   tggccgcccg  cgccgggctg  atgcaggcgc  tgccgcgcgg  cggggcgatg
36901   gtcgccgtcg  aggccaccga  ggacgaggtc  agcccgctgc  tcaccgacgg
36951   cgtcgcgatc  gccgcgatca  acggccccac  ctcgctcgtc  gtctccggcg
37001   acgagaccgc  caccctcgcc  gtcgccgccc  gactcgccga  acagggccgc
```

```
37051  cgcaccaccc  ggctgcgggt  cagccacgcc  ttccactcgc  cgctgatgga
37101  cccgatgctc  gcggagttcc  gcgcggtcgc  cgagggcctg  tcctacggcg
37151  aaccgcagat  cccggtggtc  tccaacctca  ccggcgcggt  cgccgacggc
37201  accctgctcg  gcactgccga  ctactgggtc  cggcacgtcc  gcgaggcggt
37251  ccgcttcgcc  gacggcatcc  gcgccctcac  cgacgccggc  gtcggcgcct
37301  tcctcgaact  cggcccggac  ggcacgctcg  ccgccctggc  ccagcagtcc
37351  gcccccgacg  ccgtctccgt  ccccgtcctg  cgcaaggacc  gggacgagga
37401  gcccgccgcg  gtcgccgcac  tggcccggct  gcacaccgcc  ggcgtcccgg
37451  tggactggac  ggcgttctac  gccggcaccg  gcgcccaccg  caccgacctg
37501  ccgacctacg  ccttccagta  cgagcgctac  tggcccaagg  ccacctaccg
37551  gcccgccgac  gccaccggcc  tcggcctgac  cgccgccgac  cacccgctgc
37601  tcggcgccgc  catgtccgtc  gccgggtccg  acgagctcct  gctcaccggc
37651  accctgtcgc  tcgccaccca  cccctggctc  gccgaccacg  tcgtcggcgg
37701  catggtcttc  ttccccggca  ccggcttcct  ggaactggcg  gtccgcgccg
37751  ccgaccaggt  cggctgcgac  cgggtcgagg  aactcatgct  cgccgcgccg
37801  ctgatcctgc  ccgccaccgg  caccgtccag  atgcagatcg  cggtcggcgc
37851  cgcggacgac  gacggcggcc  gcgacctgcg  cttcttcacc  cggcccgggg
37901  acgacccgga  cgccgcctgg  gcccagcacg  ccaccggccg  gatcaccgag
37951  ggcgagcgcg  tcctcgccct  cgacaccacc  acctggccgc  cccgcgacgc
38001  ggaacccgtc  gacatcgacg  gcctctacga  ccgctaccgc  gccaacggac
38051  tcgactacgg  gcccgtcttc  cgcggcctgc  gcgccgtatg  gcgccgcgac
38101  accgagatct  acgccgaggt  cgccctgccc  gaaggcaccg  ccgacgccga
38151  cgccttcggc  ctgcacccgg  ccctcttcga  cgccgtcctg  cacagcaccc
38201  tcttcgcctc  cgccgacggc  gacgaccgca  gcctcctgcc  gttcgcctgg
38251  aacggcgtgt  ccctgcacgc  cgcgggcgcg  gacgcgctgc  gcgtccggat
38301  caccagctgc  ggccccgacg  ccgtggagat  caccgccgtc  gacccgcagg
38351  gccgccccgt  cgtctccgtc  gaatcgctga  cgctgcgcgc  cgccggcccc
38401  gacgccggca  ccgccgacca  ccgtgccgac  gcgggctccc  tcttccgcat
38451  ggactggacg  ccccgcaccg  tccacgcccc  ggccacccccc  gccacctggg
38501  ccgtcctcgg  taccgacccg  atcggcctga  ccgaggcgct  caccgccgcc
38551  ggccccgaca  ccgtcacggg  actccgcgac  ggtgtcgacg  ccctcggcga
38601  actcaccgcc  ggcgacgacc  ggccggtgcc  cgacgtggtc  gcggtaccgc
38651  tgcgcggcgc  caccgaccac  gggccggccg  gtgcccacga  cctgacccgc
38701  accgtcctgg  ccctgctcca  ggaatggctg  gccgaggagc  gcttcgcccg
38751  ctccccggctg  ctgctcgtca  cccgcggcgc  ggtcgccgac  ggcgagcgcg
38801  gcccgctcga  cctggccgcc  gccccggtct  ggggcctggt  gcgctccgcc
38851  cagtccgaga  accccggccg  actgctgctc  gtcgacctcg  acgacaccgc
```

```
       38901  cgagtccgcc gcccaactcc cgttgctgcc ggccctcctg gacgccgacg
       38951  agccgcaggc cgtggtccgc gagggcaccg tccgggtcgg ccggctcgcc
       39001  cgcctggact ccggccgcgg cctcgtcccg ccgcccggca cccccctggcg
       39051  cctgggcagc cgggccaagg gcagcctcga cggcctcgcc ctgctgcccc
    5  39101  accccgaggc ccgacgaccc ctcaccggcc acgaggtccg cgtcggcatc
       39151  cgcgccgcgg gcctgaactt ccgtgacgtg ctcaacgcgt tggggatgta
       39201  tcccgggggat gcggggctgt tcggttcgga ggcggccggt gtggtcgtcg
       39251  aggtcggacc ggaggtcacc ggcctggcac ccggcgaccg ggtcatgggc
       39301  atgctcttcg gcggcttcgg accgctcggc atcgccgacg cccggctgct
   10  39351  cacccccgtc ccggccgact ggtcctggga gacgggtgcg tcggtgccgt
       39401  tggtgttcct caccgcgtac tacgccctga aggagttggg tggtctgcgg
       39451  gcgggggaga aggtgctggt gcatgccggt gccggtggtg tcggtatggc
       39501  ggcgatccag atcgcccggc atgtcggtgc cgaggtgttc gccacggcca
       39551  gtgagggcaa gtgggacgtg ctgcgctccc tgggcgtggc cgacgaccac
   15  39601  atcgcctcct cccgcaccct cgacttcgag gcggccttcg ccgaagtcgc
       39651  cggcgaccgc ggcctggacg tcgtactgaa cgcgctgtcc ggcgagttcg
       39701  tcgacgcctc gatgcggctg ctcggcgacg gcggccggtt cctggagatg
       39751  ggcaagaccg acatccgcgc cgcggactcc gttcccgacg gcctctccta
       39801  ccactccttc gacctcggca tggtcgatcc ggaacacatc cagcggatgc
   20  39851  tgctcgacct cgtcgagctg ttcgaccgcg gcgcgctggc cgcgttgccg
       39901  gtccgcagct gggacgtgcg ccgcgccggc gaggcgttcc gcttcatgag
       39951  cctggcccag cacatcggca agatcgtgct caccgtgccg caacccctcg
       40001  accccgacgg caccgtgctc ctcaccggcg gcaccggcgg cctggccggc
       40051  ctgctcgccc gccacctggt caccgagcac ggcgcccgcc acctgctgct
   25  40101  ggccggccgg cgcggccccg acgcgcccgg cgccgccgca ctccacgccg
       40151  aactgaccgc cctgggcgcc gaggtcaccg tcgccgcctg cgacgtcgcc
       40201  gaccgcaccg cgctcgccgc gctgctcgcc accgtgcccg ccgaacaccc
       40251  cctcaccgcg gtcgtgcaca ccgccggcgt cctggacgac ggcaccctca
       40301  ccgccctgaa ccccgaccgc ctcgccaccg tcctacggcc caaggtggac
   30  40351  gccgcctggc acctgcacga cctcacccgc cacctcgacc tggccgcgtt
       40401  cgtgctctac tcctccaccg ccggcgtcat gggcggaccg ggccaggcca
       40451  actacgcggc cggcaacacc ttcctcgacg cgctcgccgc ccaccgacac
       40501  gccctcggcc tgcccgccac ctcgctggcc tggggcgcct gggagcaggg
       40551  cgccggcatg accggcgcac tgaccgacca cgacctgcgc cgggtcagcg
   35  40601  acgccggcgg ccaaccgctg ctcaccgccg aacgcggcct cgccctctac
       40651  gacgccgcca ccgccgccga cgaacccctg atcgtcccgc tcggcctcac
       40701  cggcggtgcg ctgcccgccg gggtcggcgt ccccgccgtg ctgcgcggcc
```

SUBSTITUTE SHEET (RULE 26)

```
        40751  tggtccgcac cgcgggccgc cgggccaggg ccggcaccgc cggcgtctcc
        40801  cgcgccggcc tcgccgaacg cctcgccgcc ctgcccgagg aggagcgcac
        40851  ccccttcctc gtcgagctgg tgcgcaccga ggccgccacc gtcctcggcc
        40901  acggctccac cgacccggtg gacgcccgcc gcgagttccg ccaactcggc
   5    40951  ttcgactcgc tgaccgccat cgaactgcgc aaccgactcg gcaaggccac
        41001  cggcctcacc ctgcccgcca ccctcatctt cgactacccg accccccgacc
        41051  gcctcgccgt ccacctccac gacgaactcc tcggcgcgga cgccccggtg
        41101  accgtcaccg ccgccgcaca ggccgcggac ccggagcacg acccggtcgt
        41151  catcgtcggc atgagctgcc gcttccccgg cggcgtcagc tcccccgagg
   10   41201  agctgtggga cctggtggca tccggcaccg acgcgatcac cggcttcccc
        41251  gccgaccgcg catgggaccg ccacccgcag ctcgccggcg cccccggcgc
        41301  ccgcaccggc cagggcggat tcctccgcga catcgccgac ttcgacgccg
        41351  ccttcttcgg catctcgccg cgcgaggccc tggccatgga cccgcagcag
        41401  cgcatcctcc tcgaagtcgc ctgggaggcc gccgagcgcg ccggcatcga
   15   41451  cccgcagacc ctgcgcggca gcgacaccgg cgtgttcatg ggcgtcagcg
        41501  gccaggacta cgccggcctc gtgatgcgct cccgcgacga catcgccggc
        41551  cacgccacca ccggcctcgc cgtcagcgtc gtctccggcc gcctcgccta
        41601  cgcgctcggc ctggagggcc cggccctgtc cgtggacacc gcctgctcct
        41651  cctccctggt gtcgctgcac ctggccgccc aggcgctgcg cgcggggggag
   20   41701  tgcaccatgg ccctggccgg cggcgtcacc gtcatgacca ccgccgccaa
        41751  cttcaccggc ttctccccgga tgggcggcct cgcccaggac ggccgctgca
        41801  aggcgttctc cgactccgcc gacggcaccg gctggtccga gggcgccgcc
        41851  gtcctggtcc tggaacgcct ctccgacgcc cggcgcgccg gccaccgcgt
        41901  actggccgtg gtgcgcgggct cggcggtcaa ccaggacggt gcgtccaacg
   25   41951  gtctgacggc gcccaacggt cccgcccagc agcgcgtcat ccggcaggcc
        42001  ctggccaacg cgggcctgac ccccgtcgac gtggacgccg tcgaggcgca
        42051  cggcaccggc accccgctcg gcgaccccat cgaggcccag gccctgatcg
        42101  ccgcctacgg caccgaccgc gaccccgaac acccgctgct gctcggctcg
        42151  gtgaagtcca acatcggcca cacccagtcc gcggccggcg cggccgggct
   30   42201  ggtcaagatg gtcatggcca tgcgccacgg catcctgccg cagaccctgc
        42251  acctcaccga accgtcctcg cacgtggact ggtcggcggg cacggtgcgg
        42301  ctgctcaccg agcggaccgc ctggccgcgg acggatcgtc cgcgtcgggc
        42351  cggggtctcc tcgttcggca tcagcggcac caacgcccac gtcatcctgg
        42401  aacagccgcc cgccgagccc accccgccg ccgacccggg ccggccgca
   35   42451  cccaccgtgg tggcctggcc cgtctccgcg cagaccccgg ccgccctcga
        42501  cgcccaactg gaccggttgc gcaccgccgc cgccctggcg ccgctcgaca
        42551  ccgcccacac cctcgccacc ggccgctcgc tcttcgaaca ccgcgccgtc
```

```
        42601  ctgctcgcca ccgtcggcga cccggcgacc ggcgcccccg acctgcccga
        42651  ggtcgccagg ggagcggcga cgccgcaccg caccgcgttc ctcttctcgg
        42701  ggcagggtgc tcagcggtcg gggatggggc gtgaactgca tgctgctttc
        42751  ccggtgttcg cggcggcgtt cgacgaggtg gtggctgtgt tggatgcgga
    5   42801  gctgggttct gatgctgatg ggggtgtgtc gctgcgggag gtgatgtggg
        42851  gcggggggtc ggagttgttg gatcgaacgc gtttcacgca gccggcgttg
        42901  ttcgcggtgg aggtggcgtt gttccgtttg gtggcctcgt gggggtggg
        42951  gcctgagttc gtggcgggc attcggtggg tgagattgcg gcggcgcatg
        43001  tggccggggt gttctcgttg gtggatgcgt gtcgtttggt ggtggcgcgg
    10  43051  gcttcgttga tggatgcgtt gccggtgggt ggcgtgatgg ttgcggtgga
        43101  ggcggccgag gcggaggtgg tgccgctgtt ggtcgatggg gtggcgatcg
        43151  ccgcggtcaa cgggccggtc tcggtggtgg tctccggtgt ggaggcggcc
        43201  gttgggcagg tcgtggatca gttggtggag cggggccggc gggtccgtcg
        43251  gttggcggtc agtcatgctt tccactcgcc gttgatggat ccgatgttgg
    15  43301  atgccttccg ggccgtcgcc gagggcctgg agtaccacca gccgcgcatc
        43351  cccgtggtgt ccaacgtgac gggcgaggtg gccgcggcgg aggagctgtg
        43401  cgcggccgac tactgggtgc ggcacgtccg ggcgaccgtg cggttcgccg
        43451  acggcgtccg caccctggcc gagcgcggcg ccaccgcctt cctggagatc
        43501  ggccccgacg gcgtactgtc cgcgctcgcc cgcggcgtcc tgcccgccga
    20  43551  ggcgctcgtg acgcccaccc tccgcaagga ccgcgacgag gagagcgccc
        43601  tgctcgccgg actggcccgg ctgcacgtcg ccggcgtgac cgtcgactgg
        43651  agcgccgccc tgaccggcac cggcgcgcgc ggcaccgacc tgccgaccta
        43701  cgccttccaa cgcgagcggt actggccgga gttggccgcc gaacccgcgg
        43751  gcggcggcgc ggatgccgcg gacgcggagt tctgggccgc ggtggagcgc
    25  43801  gcggacgcca ccgcgctcgc cgcccacctg gacatcgacg gcgaccagct
        43851  cggcgccgtg ctgcccgcac tgtccgcctg gcgcacccgg cgccgcacca
        43901  catcggccac caacgccctg cggcaccggg agagttggga accgctgtcg
        43951  .ctcgccggca cgccgcacac cggcggcgtc ctggtgctgg tgcccgccgc
        44001  cgcgaccacc gacccctggg tcgccgacgt cgtcgccgcc ctcggcccgg
    30  44051  acgcccgccg ggtggacgtc ccggccgacg gcaccgaccg ggccgcgctc
        44101  gccgccctgc tcaccgaagc ggccgacgac accgccccga ccgccgtggt
        44151  ctccctgctc gcgctcgacg agaccagcgg cgacgacgcg gtaccggccg
        44201  gcaccaccgc caccgccgcg ctcgtccagg ccctcgccga caccggcgcc
        44251  ccggccccgc tgtgggccct gacccggggc gcggtcgccg cgctccccga
    35  44301  cgagcagccg accgcccccg cccaggccgc cgtctggggc ctcggccgga
        44351  tcgccgccct cgaactcccg cgccactggg cggactggt cgacctgccc
        44401  gccgacctcg acgagcgcac cgcacgccga ctgcccgccg cactggccga
```

```
        44451  cgccggcgac gaggaccagc tcgcgctgcg ggccaccggc gcctacggcc
        44501  gccggatcac cccggcgccc gccccgacg acgcccccgg caccgggtgg
        44551  cagccgaccg gcaccgtcct gatcaccggc ggcaccggcg cgctcggccg
        44601  gcacaccgcc cgctggctcg ccgcccacgg cgccgagcac ctgctgctgc
   5    44651  tcagccgcag cggccccgac gcgcccggcg cggccgaact caccaccgaa
        44701  ctcaccgccc tcggcgcccg cgtcaccctc gtggcctgcg acgccgccga
        44751  ccgggagcag ctgaccaggg tcctcgccga ggtaccgcgg gactgcccgc
        44801  tgaccggcgt ggtgcacacc gccggagtgc tcgacgacgg cgtgctcacc
        44851  ggcctcaccc cggaccggtt cgccacggtc ttccgcgcca aggtggcctc
  10    44901  cgccgtgctc ctcgacgagc tgacccggga caccgacctg gccgtcttcg
        44951  cgctgttctc gtccgtcgcg ggcgctgtcg gcaaccccgg ccaggccggc
        45001  tacgccgccg ccaacgcggt cctcgacgcc ctcgccgccc gccgccgggc
        45051  ccagggcctg gccggcacct cgatcgcctg gggtgcctgg gccggcgacg
        45101  gcatggcggc ccgtcacacc cgccccggcg ccgaacccgt cggcctgctc
  15    45151  gaccccgacc tcgccgtacc ggccctggcc cgcgcggtga cggagcccca
        45201  gcccaccctc gtcctcgccg acctccagca gccgcggctg ctggaatccc
        45251  tgctggcgct cgcgcccagc ccgctcctga gccggctgcc cgccgcccgc
        45301  accgcggccc gcgcggtcca ggaagcggac cgccgccgag ccggcgccgc
        45351  ggccgacctg cgcgaccaac tcgccggcac cgcacccgcc gaccgccacg
  20    45401  ccgtcctcct ccgcctggtg cggaccacgg ccgccgcggt cctcggccac
        45451  accggtgccg acgccatccg ggccgacaag cccttccgcg acctcggctt
        45501  cgactcgctc accgcggtgg aactgagcag cgctctcgcc gccgccaccg
        45551  gcctggccct cccgcccagc ctcgtcttcg accacccctc cccgcggggcg
        45601  ctcgccgacc acctgcgcgc cgaactcacc ggcgaccggc cggaatccgc
  25    45651  ccccgcggcc ccgccggcac cggtccccgc cgcggacgac gatccgatcg
        45701  tcgtcgtcgg catggcctgc cgcttccccg gcggcgtcac caccccccgag
        45751  gagttctggc agctgctcgc cgagggccgg gacggcatcg acgcgttccc
        45801  caccgaccgc ggctgggacc tcgacgtgct cggccggcga cggccagggc
        45851  cacagcgccc accgaggtcg gcggcttcct cttacgacgc ggccgccttc
  30    45901  gaccccggct tcttcgacat ctccccgcgc gaggcgctcg ccatggaccc
        45951  gcagcagcgg ctgctgctgg agaccgcctg ggaggccgtc gaacgcaccg
        46001  gcaccgaccc gacccggctg cgcggcagcc gcaccggcgt gttcgtcggc
        46051  accaacggcc aggactacgc cggcctcgtc ctgcgcgccc aggaggacgt
        46101  cgagggcac gccggcaccg gactggccgc cagcgtgatc tccggccgcc
  35    46151  tcgcctacgc cttcggcttc gagggccccg ccgtcaccgt cgacaccgcc
        46201  tgctcctcct cgctcgtcgc cctgcactgg gccgtccagg cgctgcgcgc
        46251  gggggagtgc tccctggccc tggccggcgg cgtcaccgtc atgacgacct
```

```
       46301   cgacgagctt cgccggcttc acccggcagg gcggcctggc gccggacggg
       46351   cactgcaagg cgttctccga ctccgccgac ggcaccggct ggtccgaggg
       46401   tgtgggcgtc ctcgtcgtcg aacgccgctc cgacgcgctc cgcaacggcc
       46451   atgagatcct ggccgtggtg cgcggctcgg cggtcaacca ggacggtgcg
   5   46501   tccaacggtc tgacggcgcc caacggtccc gcccagcagc gcgtcatccg
       46551   gcaggccctg gccaacgcgg gcctggcccc cggcgacgtg gacgcggtcg
       46601   aggcccacgg caccggcacc gtcctcggcg accccatcga ggcccaggcg
       46651   ctgctcgcca cctacggcca ggaccggccc gccgaccggc cgttgtggct
       46701   cggctcggtg aagtccaaca tcggccacac ccaggccgcc gccggcgccg
  10   46751   ccggcctgat gaagatggtg ctggccctcc aacacggcac gctgccgcgc
       46801   accctgcacg tcaccgagcc ctcgacccgg gtcgactggt cggccggcgc
       46851   ggtgcggctg ctcaccgagc ggaccgtctg gccgcggacg gatcgtccgc
       46901   gtcgggccgg ggtctcctcg ttcggcatca gcggcaccaa cgcccacgtc
       46951   atcctggaac agccgcccgc cgagcccacc cccacggccc ctgccgaccg
  15   47001   ccccacccgg acgcccgccg tcctcccatg ggtcgtctcg gcccgatcgg
       47051   ccaccgcgct cgacgcgcag ctcgcgcgac tgcgggcgtt cgccgccgag
       47101   cgcccggacc tgccgcccgc cgacgtcgcc cactcgctcg tcaccagccg
       47151   cgccaccttc gaacaccggg cggtcctgct ggccgcgccc gacggcatca
       47201   ccgcggccgc ccgcgccgag gcccgcgaac gcagcaccgc gttcctcttc
  20   47251   tcggggcagg gtgctcagcg gtcggggatg gggcgtgaac tgcatgctgc
       47301   tttcccggtg ttcgcggcgg cgttcgacga ggtggtggcg gtgttggatg
       47351   cggagttggc gacgggttcc ggtgggggtg tgtcgctgcg ggaggtgatg
       47401   tggggcgggg ggtcggagtt gttggatcgg acgcgtttca cgcagccggc
       47451   gttgttcgcg gtggaggtgg cgttgttccg tttggtggcc tcgtgggggg
  25   47501   tggggcctga gttcgtggcg gggcattcgg tgggtgagat tgcggcggcg
       47551   tatgtggccg gggtgttctc gttggtggat gcgtgtcgtt tggtggtggc
       47601   gcgggcttcg ttgatggatg cgttgccggt gggtggcgtg atggttgcgg
       47651   tggaggcggc cgaggcggag gtggtgccgc tgttggtcga tggggtggcg
       47701   atcgccgcgg tcaacgggcc ggtttcggtg gtggtctccg gtgtggaggc
  30   47751   ggccgttggg caggtcgtgg atcagttggt ggagcggggc cggcgggtcc
       47801   gtcggttggc ggtcagtcat gctttccact cgccgttgat ggatccgatg
       47851   ttggatgcct tccgggccgt cgccgagggc ctggagtacc accagccgcg
       47901   catccccgtg gtgtccaacg tgacgggcga ggtggccgcg gcggaggagc
       47951   tgtgcgcggc cgactactgg gtgcggcacg tccgggcgac cgtgcggttc
  35   48001   gccgacggcg tccgcaccct ggccgagcgc ggcgccaccg ccttcctgga
       48051   gatcggcccc gacggcgtac tgtccgccct ggccgcggcc tgcctgttcg
       48101   acacggacgc cgaagtggtg cccgcgctgc gcaaggggcg ccccgaggag
```

```
48151   cacaccgccc  tcaccgccgc  cgcccaactc  cacgtggccg  gcgtggacat
48201   cgactggacc  gcggtcctgg  ccggcaccgg  cgggcggcgg  atcgccctgc
48251   ccacctatgc  cttccagcgc  gagcggtact  ggccctcgct  cgccgcacag
48301   gcccccggcg  acgccggcgg  gctcggcctg  gaagccgggc  ggcacccgct
48351   gctcggggcc  gcgaccaccg  tcgccggatc  cgcggagatc  ctgctcaccg
48401   gccgcctgtc  caccaccgcc  cagccgtggc  tcgcggtcta  cgaggcggac
48451   ggccgcaccg  tcctgccggc  cgcggtcctc  gccgaactcg  ccgtccgcgc
48501   cggcgaccag  gccgactgcc  cgaccgtggc  ggaactgacc  gtcgccgcac
48551   cgctcgtcct  caccggcgcg  gcggcccagc  gcctccaggt  ccgggtggcc
48601   gcccccgacg  acaccggccg  gcgcgcgctg  tccgtgcacg  cccgacccga
48651   cgactccccc  gacagcccct  ggacgctgca  cgccaccgcg  gtcctcaccc
48701   acgacacccc  gcagcccccg  gcgccggaca  ccggctggcc  gccggagcgc
48751   gccgtgccgc  tcgacgccct  gcccaccgcc  accggcccgg  cccggatcgc
48801   ggcggcctgg  cagtggggcg  acgaactctg  cgccgagatc  gaactccccg
48851   aacccggccc  ggcggagcgg  gcattcgccc  tgcacccggc  gctgctggac
48901   accgcggtcc  gcgccggcgg  cctgctggac  ggcgacgcca  ccctggacgc
48951   cctcggctgg  cggggcctcg  ccctgcacgc  cgcgtccgcc  accgccctgc
49001   gggtccgcct  cacccggac  ggcacggaca  cctgggctct  ggaggccacc
49051   gacccgcagg  gcgctccggt  cgtctccgtc  accgggctca  ccctgggcac
49101   gcccaccgtc  gaccggtcgg  gggccggggc  ggccgatgac  ggcgcgaccc
49151   tgctcgacct  ggagtgggtg  cccgcgccgc  aggccgcgcc  caccggcggc
49201   gaccacctcc  cgtacgccgt  gctcggcgat  caactcgcgg  agctggacgg
49251   gcagttgagg  atcgcgggcg  acgggcccgg  gcgcgtcgca  tcgctggccg
49301   cgctgctgga  cggcggtgcg  ccgctgcccc  ggctcgtcct  cgcgccggtg
49351   ctgggcgtgc  cgaccgggga  aggcgacctg  cccgccgcgg  tgcgcggcac
49401   caccacggcg  gtgctggagc  tgctgcagcg  ctggaccgcc  gacgcccgca
49451   ccgccgacag  ccacctggtg  atcgtcaccc  gcggcgccgt  cgccgccggg
49501   gcggaggacg  tgcacgacct  ggcggcggcc  ccggtctggg  gcctggtccg
49551   ctcggcacag  tccgaacacc  ccggcagctt  cctgctgctc  gacctcgacc
49601   ccgccgatcc  cgcgggagcc  tcccgcgccg  ccgcgccggc  caccctggcg
49651   gccctgctcg  acgcgggcga  gacccaggcc  gcggtgcgcg  ccgacacgct
49701   caccgtcgcc  cggctgaccc  gggccgccga  cggacccgag  gccaccgccg
49751   gacacccggt  gcgggactgg  gaccgcgacg  gcaccgtcct  gatcaccggc
49801   ggcaccggcg  gcctgggcgg  cctcctggcc  cgccacctgg  tcaccggaca
49851   cggcatcaag  cacctgctgc  tcgccgggcg  ccgcggcccg  gacgcccccg
49901   gcgcgcgggc  cctgcgcgac  gaactggccg  ccctcggcgc  cgaggtgacc
49951   gtcgccgcct  gcgacgtggc  cgaccgtgcc  gcactggacc  gactcctcgc
```

```
      50001  gcaactgccg ccggagcacc cgctgaccgc cgtcgtgcac accgccggcg
      50051  tcctcgacga cgccaccgtc ggcaccctga cgcccgagcg gctggacacc
      50101  gtcctgcgcg ccaaggcgga cgccgcctgg cacctgcacg acgccacccg
      50151  cgaccgcgac ctggcagggt tcgtgctgta ctcctcggtc gccggtgtca
  5   50201  ccggcggccc cggccagggc aactacgccg ccggcaacac gttcctcgac
      50251  gcgctcgccg cgcaccgcgc cgcccagggc ctgcccggac tgtcgctggc
      50301  ctggggaccg tgggggcagg acgccggcat gaccggcacc ctcggcgccg
      50351  ccgacctggc ccgcctggag cgctccggca tgccgccgct cacccccggaa
      50401  cagggcctgg ccctgttcga cgccgccggc gcccgcggcg acgggttcgc
  10  50451  ggtggcggtg cggctcgccc gtggcgccgc cgcaccgggc gccgacgagg
      50501  tccccgcggt gctgcgtgcc ctggtgcgcg ccggcgccgc cacggcggcc
      50551  gcggccgggc acgccggtgt actggcccgc cggctggccg ccctggacgc
      50601  cgagcagcgg catcaggcgc tgctcgacct ggtccgcacc gagacggccg
      50651  cggtgctcgg ccactccggg gcggacgccg tcccggccga gcgggacttc
  15  50701  aaccggctgg gcttcgactc gctgatggcg gtcgaactgc ggacgcggct
      50751  ggccaccgcc accggagccc ggctgccggc cacgctcgtc ttcgaccacc
      50801  cgacgccgga cgcggtcgcc cggcacctcg cgtcgacgct gcccggtggg
      50851  accgcggccg gtccggaccg ttccccgctg gccgaactcg accggatcgc
      50901  cgccgagttg tcgccggagg gcgcggacga cgccacccga cagggcgtcg
  20  50951  tcggcggct gcggcacctg ctggcgcagt gggacggcac ccgacaggac
      51001  ggcggtggga cgaccgtcga cgaccgcatc gaagcggcga gcgccgaaga
      51051  ggtcctcgcc ttcatcgacc acgagctcgg ccggcaggcg gactcctgac
      51101  ccgccccact cccgtcgctc gcgcgcacca catctgagga aggtttcacg
      51151  gaccatgccg gacgaaaaga agctcgtcga ctatctgaag tgggtcacga
  25  51201  aggacctcca ccagacccgc cagcgccttc aggaggtgga ggcggggcgc
      51251  cacgaacccg tggcgatcgt cggcatggcc tgccgcttcc ccggcggtgt
      51301  gcgctccccg gaggacctgt gggagctgct gtccgcgggc cgggacggca
      51351  tcgggccgtt ccccgccgac cgcggctggg acctggcggc gctggccggc
      51401  gacgggcccg gtcgcagcgc cacccaggaa ggcgggttcc tgcccgacgc
  30  51451  ggccgccttc gacccgggct tcttcgacat ctccccgcgc gaggcgctcg
      51501  ccatggaccc gcagcagcgg ctgctgctgg agaccgcctg ggaggccgtc
      51551  gaacgctccg gcatcgaccc ggccgggctg cgcggcagcc gcaccggcgt
      51601  tttcgtcggc accaacggcc aggactacgc gcacctggtc ctcgccgcgc
      51651  aggacgacat gggcggctac gcgggcaacg gcctggccgc cagcgtgctc
  35  51701  tccggccgac tggccttcgc gctcggcctg gaaggcccgg ccgtcacccct
      51751  cgacaccgcc tgctcctcgt cactggtgac cctgcacctg gccgcacagg
      51801  ccgtgcgcgc cggcgaatgc ggcctcgccc tggccggtgg cgtcacggtc
```

```
        51851  atgacgacct cgtcgagctt cgccggcttc agcctccagg gcggcctggc
        51901  gccggacggc cgctgcaagg cgttcgccga ggcggccgac ggcaccggct
        51951  ggtccgaggg catcggcctg cttctcgtcg agcggctctc cgacgcgcag
        52001  cgcaacggcc acccggtgct cgccgtgctg cgcggctccg ccgtcaacca
     5  52051  ggacggcgcg tccaacggcc tcagcgcgcc caacggtccg tcccagcagc
        52101  gggtcatccg ccaggcgctg gccggcgccg gactcgtccc cggcgacgtg
        52151  gacgcggtcg aggcgcacgg caccggcacc cggctcggcg accccatcga
        52201  ggccggtgcg ctgctcgcca cctacggcca ggaccggccc gccgaccggc
        52251  cgttgtggct cggctcggtg aagtccaacc tcggccacac ccaggccgcc
    10  52301  gcgggcgtcg ccggcgtcat caagatggtg ctggccctgc ggcatggcgt
        52351  cctcccgcag accctgcacg tggacgcgcc ctcctcgcac gtcgactggg
        52401  agagcggcgc ggtgcggctg ctcaccgcac ccgtcgcctg gtccgagggc
        52451  gacgaccggg tgcgccgggc cggcgtctcg tcgttcggca tcagcggcac
        52501  caacgcccac gtcatcctcg aacaagcccc cgatcagccg gaaccgaccg
    15  52551  cggaagagac ggctgccgcg gcgcccggcg gcaccgccga ggagcgggcc
        52601  gccgctcccg tcgccccgcg cgccgtgccg tggccggtcg cggcacgcac
        52651  cgccggcgcc ctcgacgccc aactggtccg ggtccgcgcg ctgaccaccg
        52701  cgcccggccg caccgccgcc gacgtcggtc acgcgctggc caccgcccgt
        52751  accccttcg agcaccgggc gctgctggtc cacgagggcg gcgccgtcac
    20  52801  cgaggtggcg cgcggcgccg tccccaccgg tgaccggggc gggctggccg
        52851  tgctgttctc cggacagggc tcccaacggc cgggcatggg gcgcgaactc
        52901  cacgcccgct acccggtctt cgccgccgcc ttcgacgaga ccgtcgccct
        52951  gctcgacgcc cggctcggca cgtcgctgcg cgacatcgtc tgggaccagg
        53001  accgcacccg gctcgacgac acccgccaca cccagcccgc gctgttcgcc
    25  53051  gtcgaggtcg cgctgtaccg cctgctggcc tcctggggca tccggcccga
        53101  ccacgtcacc ggacactcca tcggcgagat caccgcggcg cacgtcgccg
        53151  gtgtgctgac cctcgcggac gcctgcaccc tggtggccgc ccgcgccacc
        53201  gccatgagcg aactgccgcc cggcggcgcc atggtggcgc tggaggccac
        53251  cgaggacgag gtgcgtccgc tgctcaccga cgacctcgcg atcgccgcgg
    30  53301  tcaacgcccc ccggtccgtg gtcgtcgccg gcgccgagga cgccgccctc
        53351  gccgtccgcc ggcacttcga cgacctgggc cgccggacca cccggctccc
        53401  ggtcagccac gccttccact cgccgctcat ggacccgatg ctcgacgcct
        53451  tccggacggc cctcgccccg ctgaccttcg ccgagccgga gatcccggtc
        53501  gtctccaacc tcaccggcct cccggccacc gccgaggaac tcgccacccc
    35  53551  gcactactgg gtgtgccacg tccggcaggc cgtccgcttc ggcgacggcg
        53601  tgcgcgccct cgccgaccgc ggcgtgcgga ccttcctcga actcggcccg
        53651  gacggcgtgc tgtccgccct ggtccgggag aacctccccg agccgggcct
```

```
      53701  ggtcgccgtg  cccgtgctgc  gcaaggagcg  gcccgaggag  accaccgtgc
      53751  tggccgccct  gggaaccctg  tgggcgcacg  gcgcggacgt  ggactgggac
      53801  gcggtgttcg  ccggcacccg  caccccgcag  gccgaccccg  tcgagctgcc
      53851  gacgtacgcc  ttccaacgcg  cccgctactg  gcccacccct  ggcgcccgcc
  5   53901  acggcgaccc  ggccgacctc  gggcagaccg  ccgccgccca  cccgctgctg
      53951  ggcgccgccg  tcaccctcgc  cgacgccgac  gagaccgtgc  tcaccggccg
      54001  cctcgcgctg  ccctcccacc  cctggctcgg  cgaccaccgc  agcgacggcc
      54051  ggatcaccgt  ccccggcgtc  gccttcgccg  aactcgccgt  ccgcgccggc
      54101  gacctgagcg  gcaccccgca  cctggcgcgg  ctcgacctgc  cggcgccgct
  10  54151  caccctcggc  gacggcgaca  ccgtcaccct  ccaggtccgg  gtcggcgccc
      54201  ccgaccccgc  gggccaccgg  ccgctgaccg  tccacgcccg  cctcgcagcc
      54251  accgaggacg  ccccctggac  cacctgcgcg  accggtctgc  tcgccccgga
      54301  cgcccccgaa  gcgcccgcgg  atccgatcgg  cccggccgac  gccgggtggc
      54351  cgccgcggga  cgcccgcccg  gtgcccgtcg  ccgacctcga  cgcggccgcc
  15  54401  accgccgcag  gccgccacta  cggcccccat  ttccagggcc  tgaccgggct
      54451  ctggcggcgc  gacggcgagg  tcttcgccga  ggtggccctg  cccaccgcca
      54501  ccgccgccga  ccgcgccttc  ggcatccacc  ccgcgctgct  ggccaccgcg
      54551  ctccgcgcca  ccgccgcact  ggacgacgac  cacaccgccg  gccacacccc
      54601  cgaaccgacc  ggcatcaccg  gactcgccct  gcacgccacc  ggggccaccg
  20  54651  cactgcgggt  ccggctcacc  gcgaccgggc  ccgacaccgt  ggccctcgcc
      54701  gccgcggacg  ccacgggcgg  cgcggtcctg  accgccgaca  ccgtcaccct
      54751  cggctccccg  caggaccgcc  cggctcccgc  accggccggc  cacaccgggc
      54801  agggcggcct  gttccacctc  gactgggtgc  cggtcgaccc  cggcagccga
      54851  gccaccggca  cccgctgggc  cgtcgtcggc  gacgacgaac  tcgacctcgg
  25  54901  ctacgccctg  caccgcgccg  acgagacggt  cagtgcctac  gcggcgtcgc
      54951  tgggcggagc  catcggcgac  agcggtctgg  cgcccgacgt  cttcctcgtc
      55001  cccgtcgtcg  gcggcccgga  cgccgggccc  gacgcggtgc  acgccgtcac
      55051  cgcccgcgcc  ctggggctgc  tccaggagtg  gctgaacgag  ccgcggttgg
      55101  ccggcgcccg  cctggtcttc  gtcacccgcg  gcgccgtcgc  ggtgcccggc
  30  55151  gagaccgtca  ccgacccggc  cggcgccgcc  gtctggggcc  tgctgcgctc
      55201  cgcccagacc  gagaacccgg  gcagtctgct  gctggtcgac  ctcgacgacg
      55251  cgttccggtc  cgccgggatg  ctgccgcacg  tcctcaccct  cgacgaacag
      55301  cagctcgtcg  tccgcgacca  cgcggtccgc  gccgcccgcc  tggcccggct
      55351  gccggagccg  gccgccggca  ccgcgccggc  ccgcgcctgg  gacccggacg
  35  55401  gcaccgtcct  gatcaccggc  ggcaccggcg  gcctgggcgc  cgcgctcgcc
      55451  cgccacctgg  tcaccgtccg  cggcgcccgc  cacctgctgc  tcgccggccg
      55501  ccgcggcccc  gaggcgccgg  gcgccggcga  actggtggcg  gagctgaccg
```

```
  55551  cacagggcgc ggacgtgcgg gtggccgcct gcgacgtcgg cgaccgcacc
  55601  gccctcgacg cgctcctcgc cacggtcccc gcggcgcacc cgctgaccgc
  55651  cgtcgtgcac accgccggcg tcctggacga cgccctgatc ggctcgctca
  55701  cccccgacca actggccacc gtgctacggc ccaaggccga cgccgcctgg
5  55751  catctgcacg acgccacccg cggcctcgac ctggccggct tcgtcctgta
  55801  ctcctcggtc tccggcgtcc tgggcagccc cggccagggc aactacgccg
  55851  ccgccaacgc ctacctcgac gcgctcgccc ggcaccgcgc cgaccagggc
  55901  ctcccggcgc tctccctcgc ctggggcccc tggggtcggg gcagcggcat
  55951  gaccgcgtcg gtcagcgacg ccgacctgga gcggatggcg cgcggcggcc
10  56001  tgccgccgct gaccgtcgag gacggcctgg ccctgttcga cgccgccgtc
  56051  ggccgccccg agccggccct ggtgcccagc cgcatcaacg tcgccggcct
  56101  gcgggaccag caggcactgc cggcactctg gcgcgacctg gtaccgcggg
  56151  cccgccgcac cgcggccacc gccgaccgct ccccggtcac ggtgcgcgag
  56201  cgcctccgcc acctcgacga gaccggccag gagcagctgc tcatcgacct
15  56251  cgtcgtcggc tacaccgccg gcctgctcgg ccaccccgac cccaccgccg
  56301  tcgatcccga acggggcttc ctggagctgg gcttcgactc cctggtctcg
  56351  gtcggcctgc gcaaccagct cgccgagatc ctcggcctgc gcctgccctc
  56401  gtccatcgtc ttcgacagca agtcgccggt gaagctggcg cgttggctgc
  56451  accaggaact cgccaacggc ccccagccgg gcgccaccgg ccccgccgcc
20  56501  gcggacgccc gtcccgcggt gcgctccgac gacaccctgg agggcctgtt
  56551  ctacaacgcg gtgcgcggcg gcaagctcgt cgaggcgatg cggatgctca
  56601  aggccgtcgc caacacccgg ccgatgttcg acacccccgc cgagctggag
  56651  gagctctccg agccggtgac gctcgccgac ggcccgggcc ggccccggct
  56701  gatcttcgtc agcgccccgg gcgccaccgg cggcgtccac cagtacgcgc
25  56751  gcatcgccgc gcacttccgc ggcagccgcc atgtctccgc gctgcccctg
  56801  atgggcttcg cccccggcga gctcctcccg gccaccagcg aggccgcggc
  56851  ccgtatcgtc gccgagagcg tcctgatggc cagcgagggc gaaccgttcg
  56901  tcatggtcgg ccactccacc ggcggctcgc tggcctacct cgccgccggc
  56951  gtcctggagg acacctggga cgtccggccc gaagcggtgg tcctcctcga
30  57001  caccgcgtcc atccgctaca accccggcga gggcaacgac ctggaccgca
  57051  ccacgaggtt ctacctggcc gacatcgact cgccctcggt gacgctcaac
  57101  agcgcccgga tgtccgccat ggcccactgg ttcatggcga tgaccgacat
  57151  ccaggcgccc gcaccgaccg ccccacccct cctcgtgcgc gccgcccggg
  57201  ccctcgacgg cttccggctc gacacctcgt ccgtccccgc cgacgaggtc
35  57251  cgggacatcg acgccgacca cctctccctc gccaaggagc actcggcact
  57301  gaccgcgcag gccatcgagg gatggctcgc ggaactgccg gaccccgcgg
  57351  cctgatctcc ggcccggccg gcccccgaca ggccgggccc atcccccccac
```

```
      57401  cggggcggcg ccacccaggt gcgccccggc gcacccggcc ccgcgcaggc
      57451  agccggagcg cccatccccc cgcatggcac caccctccag aggagtcctt
      57501  ccatgagcac accgaccgca ccgccctccc tgaaagcgga ggtgccgccc
      57551  gtcctgcgcc tgagcccgct gctgcgcgaa ctccagtccc gcgcccccgt
   5  57601  ctgcaaggtc cgcacccccg ccggcgacga gggctggctg gtgacccggc
      57651  acaccgaact caagcagctg ctgcacgacg accggctggc ccgcgcccac
      57701  gccgacccgg ccaacgcccc gcgttatgtg cacaacccgt tcctggacct
      57751  gctcgtcgtc gacgacttcg acctggcccg cacgctgcac gccgagatgc
      57801  gctccttgtt caccccgcag ttctcggccc gccgcgtcat ggacctgacg
  10  57851  ccgagggtgg aagccctcgc cgagggggta ctggcccact tcgtcgccca
      57901  gggaccgccc gccgacctgc acaacgactt ctcgctgccg ttctccctgt
      57951  cggtgctgtg cgcgctcatc ggcgtcccgg ccgaggaaca ggggaagctg
      58001  atcgccgccc tcaccaaact gggcgaactc gacgacccgg cacgcgtcca
      58051  ggaaggccag gacgagctgt tcggcctgct gtccggcctg gcacgccgca
  15  58101  agcgcatcac acccgaggac gacgtcatct cccggctctg cctgaaggtg
      58151  ccctccgacg agcgcatcgg cccgatcgcc tccggtctgc tcttcgccgg
      58201  cctggacagc gtcgccagcc acatcgacct gggcacggtg ctgttcatcc
      58251  agcacccgga ccagctcgcc gcggccctgg ccgacgagaa gctgatgcgc
      58301  ggcgccgtcg aggagatcct gcggtccgcc aaggccggcg gttcggtgct
  20  58351  cccgcggtac gcgaccgccg atgtaccgat cggcgacgtg accatcaggg
      58401  ccggcgacct ggtgctgctg gacttcaccc tggtgaactt cgaccgcacg
      58451  gtcttcgacg agccggagct cttcgacatc cggcgcgccc ccaacccgca
      58501  cctgacgttc ggccacggca tgtggcactg catcggcgcg ccgctggccc
      58551  gggtcaatct gcgcaccgcc tacaccctgc tgttcacccg cctgcccggc
  25  58601  ctgcggctgg tgcgccggt cgaggaactg cgggtgctgt cggggcagtt
      58651  gtcggccggc ctgacggagc tgcccgtcac ctggtgacgt gatgtcggac
      58701  ccgcccggcc cggttccggg caggcggaag tgagggcccc cgccccgctc
      58751  ggccggggggc cctcacgcac gggggagcgg gctcctcagt ccgccgcgac
      58801  gctgatcgcc ccggacgggc agagcgccgc cgcgtcgcgt acgtcgccgg
  30  58851  gatccgcggc gtccgccgcc ccggccagga cggtcaccag accgtcgtcg
      58901  tcctggtcga acaggtccgg tgcggtcagg acgcactggc cggccccgac
      58951  acagcggccg gggtccacgg tgatgcgcac gatcgttcct ccgagggtcg
      59001  gttcgtcgcc ggggtcgggc cgccgccggg cggcccgacc gggctcactt
      59051  ccaggtgacg ggcagttcgt gcaggccgaa cagcaccccg tcgtacttca
  35  59101  acggcagttc ctccaccggc acggcgagcg tgagggaggg gatccgggcg
      59151  aagagcgtgc ggtaggcgat gtccatctcc tcccgcacca ggttctggcc
      59201  caggcactgg tggacaccgt agccgaacgc cacgtgcgag cgcgccgagc
```

```
       59251   gctcggggtc gaactccgag ggcgccgcga aggccgtggc gtcgtggttg
       59301   gcggcggcga tcagcggcac gatgccctcg ccggcccgga tcagctgacc
       59351   gccgatctcc acgtcctcga cggccacccg gaatgccacc agatcggcca
       59401   ccgagtagta gcgcagcagt tcctcgacga tccggtcgtc gccgatccac
 5     59451   cgcgggttgg tgagcaactg caccacgccc aggccgatgt tgttcgccgt
       59501   cgtctcgtgc ccggcgatca gcagcagcat cgccaccccg gacagctcct
       59551   gccgggacag tgtgcccgcg gcgatcagcc ggctgatcag gtcgtcaccg
       59601   ggtcgccgct gcttgatctg gatgagtcgg cccagatagc gcagcagcgc
       59651   ctgggtcgcc ttgtcgcgct cctcgtcggt cgaactgagc cggaccagca
 10    59701   cccgggtccg gtcctcgaag aagtcccggt cgaccttggg caccccgagc
       59751   agactggaga tcaccaggga cggcaccggc agcgcgaagg actcgacgag
       59801   gtccgccgag gtgcccgcgg ccagcatcgc gtcgatccgc tcgtccacga
       59851   tctgctggat caccggccgc agttcccgca ccttgcgcac ggtgaactcg
       59901   gggatcaagg tcttgcggaa ccggccgtgc tccggtgggt ccatcgccac
 15    59951   gaaccacccc ggcacctcgt actgcgacgg ggccccaccg gtccgaccgg
       60001   ccttggggaa accgggtttg gaggggtcgg cgctgatccg ggggtcggtc
       60051   agcaccgccc ggacgtcctc gtgccgggtc accaaccaga ccgggccgct
       60101   ggggagttcg gagcggacca ggcccgcccc gccgcggtag gtcgcgtact
       60151   ccggcggggg gaagggccgg cccggcctgc gcagcgggaa ggccaccggg
 20    60201   cactgcggcg cggcggtccg cgcgtcggct tcggtgctca tgccactccg
       60251   tagaacgcgc ggatccggcc ggtgatgaac tcctgctcct gtgcggtcag
       60301   tccggtgtgc gtcggcagat agaagccgtc ggcgctgagc cgggccgcgt
       60351   tcaacgcggg ccagtcggcg gagaagtaac cgggctgacg gctcatcggc
       60401   ttgaagaaga cccgggtctc gatgccctcg ccggcgaggt acgcgcacag
 25    60451   ctcgtcccgc cgctcggccc gcaggtcgta catccacagc acgtcgcgcg
       60501   gcggcatcag cgtgatgccg gggatgtcgc gcagcgcctc gtcgtagcgc
       60551   ttctcgatgt cgcggcgcag ggcgaggatg gtgtccagct gctcggtctg
       60601   cgcgagcgcc accgcggcct gcatgttggt catccggaag ttgtaggcca
       60651   gcttcttgtg caggaagctg tggtccttgg tgaacgccat cgcccgcaga
 30    60701   tgggccatct gctcggccag gtgcgggtcg tgggtcaggc agacgccgcc
       60751   ctcgccggcc gagatgatct tgttggcgaa gagcgagaaa caggcgatgt
       60801   cgccccgcgg ccgcaccccg tgcgcctcgg cggagtcctc caccacccgc
       60851   aggttgtact cgtacgccag gttcagcacg gcgtccatgt cgcactgccg
       60901   gccgtagatg tgcaccggca tgatcacttt ggtgcgcggg gtgatcttct
 35    60951   cctcgatgcg ggacacgtcg atgttcaggt cgtcgccgca gtccacgaac
       61001   accggcgtgg caccggtgta ggtcaccgcc caggcggacg cgatcatcgt
       61051   gaactccggg acgatcacct cgtcaccggg ccgacgccc agcgcgcgca
```

	61101	gcgccagcgt	cagcgccgtg	gtgccggagg	agcaggcgac	gccgaacggc
	61151	acgtcgttgt	acgcggcgaa	cgcctcctcg	aaccgcctga	cgtacggccc
	61201	ctgcgaagag	atccagccgc	cgccgacggc	ctccgtcaca	tagtcgagct
	61251	cgcggccctg	gagccacggc	atggacaccg	gatacgtaaa	ggacatgggt
5	61301	tttgagtctt	cctcggtcag	tcggttgcca	gggcgggaag	gccgagcagc
	61351	aggtccgccg	ccgcggcccg	gccgccggcg	tcccgcagca	gcccggcgaa
	61401	gtgctccgcg	cgctcggtga	aggagggttg	gtcgagcacg	cgggtgatct
	61451	tgtccaggac	gtcctcggtg	tccacggtct	ccggccggtc	cagggtcagg
	61501	ctcaccccga	agtcctggcc	ccggatcgcc	tggtcgtcgc	agtccaccca
10	61551	caacggccgg	accaccagcg	gctttccgaa	gtacaggccc	tcgtggtagc
	61601	cgttgccacc	ggcatgggtg	aagaacgcct	tcacgttcgg	atgggccagc
	61651	acgtccagct	gcgacggcac	ccagccctcg	atccgcaggt	tgtccggcag
	61701	ctcggcggcc	ggcggcagca	actcctgttg	gccgcgcggg	agtttccaca
	61751	acacctggtg	gccccggccg	tccagtcgcc	gggcgacctc	caccagcgac
15	61801	gccacctgct	cacgggtcag	ccgggtgatc	gtgccgaagc	ccatgtacac
	61851	cacggacttc	tgcgccgaca	gccagtccga	caggccgtcg	tcgtccggtg
	61901	cctggggcag	cggcggcacc	atcgtgccca	ccagccgcag	cttcggatgc
	61951	atcgggaacg	ggtagtccaa	ctcccttacg	gagtagcaca	agacctgctc
	62001	cgcatggtcg	atccgcgcca	tcatctgccg	cgcctggggc	gcgatgccca
20	62051	gctcggtgcg	gacccggttg	tcctcctcga	cgaccttgcg	gacgtccgac
	62101	gtcaggaaca	tcccgagcgt	ccgcagccgg	aacagctggt	tctcgatccg
	62151	ctgagccagg	gacatcgcgg	ccggcagccc	cgagtgcggc	accgggaaac
	62201	ccgacggggt	gtaggacttg	gcgaacggga	cgtgcgaggt	gaggacgttg
	62251	ctcggcacga	acggcacccc	gagcacgaac	ggaatgccct	tggtgatcgc
25	62301	cagctcgtac	ccgaactggc	acatgctctc	gatcaccatc	agcgccggct
	62351	cgacctcctc	gacgatctcc	tccaggcggc	ggtacttcgc	catccgcgac
	62401	tccggcgcga	acgaatgccg	aatcaccgcc	gcgtgcgcct	tgaaccgcga
	62451	ccgctgcgtc	acctccgcat	acgtcgcgtc	gtcccatgtg	accgccgaca
	62501	tctgcgagac	ggtgtcgccg	agcgacgcga	accgaaccgg	ctgccgtcc
30	62551	accacggccg	ccacctcgtc	gcgcgctttc	tcgtcggtgg	cgaaccacag
	62601	gtccgccacg	tcgcgccggg	acaattcccc	ggccagcacg	agcagcggat
	62651	tgagcaggcc	gctttcggca	taactgacga	acaggatcgg	ccgccgattc
	62701	gcgcccatgg	acaacacccc	tcggaatgtg	gcgggccgcc	gggcccgcgc
	62751	gccacgcacc	cgcccggccc	ggtcgccggg	tgagtgcatt	cgccgacgcc
35	62801	gccacccgag	gcgcgtgttg	ccggaaggaa	gggtcaccgg	ccggcacccg
	62851	gaacgcgccg	cgtggaaaac	gggtcggtta	cttggtctca	tgccacggac
	62901	cggggaatca	ctagtcttcg	gcgcgcgacg	gccctttccg	ggccgtgtgg

```
       62951   ccaatgcccg tccccggcgc ccgtcattcc ttagggaaaa gtacagcgtt
       63001   tgcgaacgta cgatccggca cgcagaggtg acctgaggcc aactttccg
       63051   cagggglgag caaggcatga cgatcggagc cgacgaggac ccggtggtgg
       63101   tcgtcggaat ggcctgccgt tatccgggtg gggtcgccgg cccggaggac
  5    63151   ctgtgggaac tggtccgcac cggccgcgac gcgaccaccg ccttcccgga
       63201   cgaccgcggc tgggacctgg ccgcactggc cggcgacgga cccggccgca
       63251   gcgcgacccg cgagggcgga ttcctcaccg gcgccgccga cttcgacgcc
       63301   gccttcttcg gcatgtcgcc ccgcgaggcc gtctccaccg acccgcaaca
       63351   gcgcctcgtc ctggagaccg cctgggaagc cctggagcgc gccggcatcg
  10   63401   acccgcactc cctgcgcggc agccgcaccg gggtcttcgt cggcgccagc
       63451   ggccaggact acgccgccgt cacccacgcc tcgcccgacg acctggacgg
       63501   acacgccctc accggcctgg cccccggcgt cgcctccggt cgcctggcgt
       63551   acgtcctggg cctcgaaggc cccgccgtca ccgtcgacac cacgtcctcc
       63601   tcgtcgctgg tcgcgctgca ctgggcggtc cgcgccctgc gcgcggggga
  15   63651   gtgcagcacc gccctggccg gcggcgtcac ggtgatgtcc accccggccg
       63701   ccttcgtcgg ccacacccga cagggcggcc tcgcgcccga cggccgctgc
       63751   aagccgttct ccgacgacgc cgacggcacc gcctgggcgg agggcgtcgg
       63801   catcgtcgtc ctggagcacc tgtccaccgc ccgcgccgcc ggcaaccccg
       63851   tcctcgccgt gctgcgcggc tcggccgtca accaggacgg cgcctccgac
  20   63901   ggcctcaccg cacccagcgg tcccgcccag gaacgcgtca tccgcgccgc
       63951   cctcgccgac gcccgactcg cccccgccga catcgatctc gtcgaggcgc
       64001   acggcaccgg cacccggctc ggcgacccca tcgaggcccg ggcgctgctc
       64051   gccgcctacg gccaggaccg ggacccggac cgaccgctgc gcctcggttc
       64101   cctgaagtcc accctcggcc acgcacaggc cgccgccggc atcggcggag
  25   64151   tgatcaagac cgtcctgacc ctgcggcacg gcctgatgcc gcgcatccgg
       64201   cacctggcca cccccacccg ccaagtcgac tggtcccagg gcgccgtggc
       64251   cccctcacc gaccacacgc cctggccacc ggccgaccga ccgcgccgcg
       64301   ccggcgtctc ctccttcggc atcagcggca ccaacgccca tgtgatcctc
       64351   gaagaggcgc cgcccgccga cgtccccgtc acccggcccg gcaccctccg
  30   64401   ccccagcacc gtcccctggc cggtctccgc cgccacgccc gaagccctcg
       64451   acgcccaact cgcccggctc cgcgcccacc tgcgcacccca ctcggacctg
       64501   gacccgctgg acgtcggcta ctccctggcc accggccgcg ccgcgctccg
       64551   ccaccgggcg gtcctcctgc cgcccgccga cggcaccgcc gcggacgccg
       64601   tcgagcacgc ccgcggtgcg gcccaccagc gccgcaccgc cgtcctcttc
  35   64651   tccggccagg gcagccagcg cccgggcatg ggccgcgaac tcgccgcccg
       64701   cttcccggtg ttcgccgacg cactggacga cgcgctgcgc gccctggacc
       64751   ggcacctgga cggcccggtg cgcgaggtga tgtggggcac cgacgccgcg
```

```
       64801  ctcctggacc ggaccggctg gacccagccc gccctgttcg ccgtcgaggt
       64851  cgccctccac cgcctggtcg cgtccctcgg cgtcacccce gacttcgtcg
       64901  gcggccactc cgtcggcgag atcgccgccg cccacgtcgc cggcgtcctg
       64951  agcctggagg acgcctgccg cctggtggcc gcccgcgcca cgctgatgca
   5   65001  ggcgctcccg gccggcggcg cgatggccgc gctggaggcc accgaggacg
       65051  aagtggcccc gctgctcggc gcacacctcg cgctggccgc cgtcaacggc
       65101  cccaccgcgg tcgtcgtcgc cggagccgag gacgccgtgc ggcaactgac
       65151  cgcccgcttc gccgaccgcg gccggcgcac cagccggctg gccgtctcgc
       65201  acgccttcca ctcgccgctg atggagccca tgctcgacgc cttccgggac
   10  65251  gtcgtgagcc gactgacctt ccaccagccg tcgatcccgc tggtctccaa
       65301  cctcaccggt gaactcgccg gcagtgagat caccagcgcc gagtactggg
       65351  tccggcacgt ccgcgacacc gtccgcttcg ccgacggcat caccgcactg
       65401  gccaaggccg gcgccgacgt cctgatcgaa ctcggccccg gcggcgtgct
       65451  gtccgcgatg gcccgcgaca ccctcggccc cgacagcacc accgacgtcg
   15  65501  tccccgccct gagcaaggga cggcccgagg agaccgcctt cgccggcgcc
       65551  ctcggccgcc tgcacaccct cggcgtcccc gtcgactggc ccgccttcta
       65601  cgccggcacc ggcgcccgcc gcgtcgaact gcccacctac gccttccagc
       65651  acgtcgcccg ctggcccacc ccgccccgcc cgaacggcgc cgggcccggc
       65701  gccctcggcc acccctgct cggctccgcc gtcgaactcg ccgacggcgg
   20  65751  cggcaccgtc tgctccggcg ccctctccct ccgcacccac ccctggctcg
       65801  ccgaccacac cgtcgccggg cgggtcgtgc tgccggccac cgcgctgctg
       65851  gaactcgccg tgcgcgccgg cgacgaggcg ggctgcgacg tcctgcacga
       65901  actccacctc accacccgc cggccctgcc cgacgacgcc gccctgcacg
       65951  tccaggtgca cgtcggcccc gccgacacca ccgggcgccg cgccgtcacc
   25  66001  gtccacaccc gccccgacca ccaccggcc ggcgactgga cccgatgcgc
       66051  caccggcacc ctcggcagca ccccgccgtc cgcagccgaa gccgccacgg
       66101  gcggcacccc ggccgcctgg ccgccggccg acgccgaacc cctcgacctc
       66151  gccgaccact acgaacggct cgccgaccgc ggcttcgact acggcccgac
       66201  cttccgcggc ctgcgggccg cctggcgacg cggcgcggag atcttcgcgg
   30  66251  acgtggaatg cccgcccggc accgccgacg acgcccccga ccacggactg
       66301  caccccgccc tgctcgacgc ggcccggcac gccgccatgg cggtggacgg
       66351  caccgtgccc gtcgcctggc acggcgtccg gctgcacgcc gtcggcgcca
       66401  ccgcgctgcg ggtccgcatc cgccccacca cgaccggcac gctgaccctc
       66451  accgcggtcg acgtgcacgg cgcgccggtc gtcaccgtcg aggccctcac
   35  66501  cgcccgcccg ctgaccgacg aggaacgcgc cgccccgcgg acgccgcggc
       66551  aggcccgcgg cgagacgccc gccgacgccc gcccggcccg gcccgcggcg
       66601  gcccgccccg gcccggccgg cgaacccctc ccggacacca ccgggtccca
```

```
       66651   ccccaccgcc  ggccacctcg  ccgcgctgcc  gccggccgcc  cgggagcgcc
       66701   agctgctgga  cctggtgcgc  acccaggccg  ccgccgtcct  gggccacccc
       66751   ggccccgagg  ccgtcggcac  ccgcagcgtc  ttcaaggagc  tgggcttcga
       66801   ctcgttggcc  ggcgtcgaac  tcgccgaccg  gctcaccgcc  cgcaccggac
   5   66851   tgcgcctgcc  ggccaccctc  gtcttcaact  tccccacccc  cgaacgtgct
       66901   gcccaccgcc  tcggggaact  cctcgccgca  accgcccccc  tcgaccccgg
       66951   ggcgtacgga  gaggaactca  ccaggttcga  ggcgatcgtg  acgaacctgc
       67001   cgcaggacgg  ccccgaacgc  cgggccgtcg  cggaccggtt  ggacgccatc
       67051   gtctccgcac  tccgccagaa  ctcgcctgca  gaggtgccct  cctcggacga
  10   67101   ggacatcgac  acggtgtcgg  tcgacagact  gctcgacatc  atcgatgaag
       67151   agttcgaaac  cacatagaga  aattgttgct  ttcgttcgcg  acccgatgac
       67201   gaggacggac  cgatgcagga  accccagcaa  ggccagccgg  accagcagga
       67251   gaaaatcgtc  gactatctcc  ggcgggtcac  ttcagatctt  cgccgtgccc
       67301   gccgccgcat  tggcgaactg  gaatccaagg  acaacgagcc  catcgccatc
  15   67351   gtcggaatgg  gctgccgact  tcccggcggc  gtcaattcgc  cggaatccct
       67401   gtgggacctg  gtgcgttccg  gcggcgacgc  catttccgga  ttccccgtcg
       67451   accgcggctg  ggacctggaa  accctcaccg  gaaacggcga  cggcagcagc
       67501   gccacccacg  aaggcggatt  cctctacgac  gccgcggaat  tcgacgccgc
       67551   cttcttcggc  atctcgccgc  gcgaggcgac  tgctatggac  ccccagcagc
  20   67601   gcctcctcct  cgaagtcgcc  tgggaggcgc  tggagcgcgc  cggcatcgcc
       67651   cccacagccc  tgcgcggcag  ccggtccggc  gtgttcgtcg  gctcctacca
       67701   ctggggcgcg  ccctcggccg  acgccgccac  cgaactgcac  ggccacgccc
       67751   tgaccggcac  cgccgccagc  gtgctgtccg  gccgcctggc  ctacaccctc
       67801   ggcctcgaag  gcccggccgt  caccgtcgac  accgcctgct  cctcctccct
  25   67851   cgtcgccctg  cacctggcgg  cccagtccct  gcgcgtcggc  gaatcctcgc
       67901   tcgccgtgat  cggcggcgtc  acgatcctca  ccgagccgtc  cgtcttcgtc
       67951   gagttcagcg  cccagggcgg  cctggcaccg  gacggccgct  gcaaggcgtt
       68001   ctccgacgcc  gccgacggca  ccggttgggc  cgagggcgtc  ggtgtcctcg
       68051   tcgccgagcg  gctctccgac  gcgcagcgca  acggccatcc  ggtcctcgcc
  30   68101   gtgctgcgcg  gctccgccgt  caaccaggac  ggcgcctcca  acggcctgac
       68151   cgcccccaac  ggcccctccc  aggaacgggt  catccagcag  gccctcgccc
       68201   ggaccggcct  gaccccccgcc  gacatcgacg  ccgtcgaggc  gcacggcacc
       68251   ggcacccggc  tcggcgaccc  catcgaggcc  caggccctgc  tcgccaccta
       68301   cggccaggga  cacaccccccg  accagccgct  gtggctcggc  tccctgaagt
  35   68351   ccaacatcgg  gcacacccag  gcggccgccg  gcgtcgccgg  tgtcatcaag
       68401   atggtcatgg  cgctgcgcca  cggccacctg  ccgccgaccc  tgcacgccga
       68451   cgcgccctcc  tcgcacgtgg  actggtccgc  cggatcggta  cgcctgctga
```

SUBSTITUTE SHEET (RULE 26)

```
        68501   ccgagggcca gcagtggccg gagaccggac gtccgcgccg ggccgcggtg
        68551   tcctcgttcg gcatcagcgg caccaacgcg cacgccctgc tggaacaggc
        68601   accccacccc gcggacaccg cggacgccgg cgacgacgcc gcgcccaccg
        68651   aaccggccgg cgcgcccgcc gcgctgccct ggatcgtctc cggacactcc
    5   68701   ccgcaggcgc tgcgcgacca ggccgccgcc ctggccgcca gggtcgagac
        68751   cgaccccgcg ctccgccccc aggacatcgg gcacaccctg cacaccgccc
        68801   gcgccctgct cgaacgacgc gccgtcgtcg tcgcccccga ccgcgccgaa
        68851   ctcctcgcgg ctacccacga gttggccgcc ggccggtccg cgaacgccgt
        68901   cgtcgagggc ctcgcggacg tcgagggtcg gacggtgttc gtgttccccg
    10  68951   gtcagggttc gcagtgggtg gggatggggg cccaactcct cgatgagtcg
        69001   gcggtgttcg cggagcggat tgccgagtgt gcggcggcac tcgccgagtt
        69051   caccgactgg tcgctggtcg atgtgctgcg gggtgtggtg ggtgcgccgt
        69101   cgttggagcg ggtcgatgtg gtgcagccgg cgtcgttcgc ggtgatggtg
        69151   tcgttggctg cgttgtgggg ttcccgtggt gtgttgccgg atgcggtggt
    15  69201   ggggcattcg caggtgaga tcgctgccgc ggtggtgtcg ggtgcgctgt
        69251   cgttgcggga cggggcgcgg gtggtggcgc tgcggagtca ggccattggt
        69301   cgtgcgttgg cggggcgggg cgggatgatg tccgtcgcgc tgtcggtgga
        69351   cgtgctcgaa ccgcggttgg tcgagttcga ggggcgggtg tcggtggccg
        69401   ccgtcaacgg cccgcgctcc gtcgtggtcg ccggcgagcc cgaggcgctg
    20  69451   gacgcgctgc acgcccggct gaccgccgac gacatccggg cccgccggat
        69501   cgcggtggac tacgcctcgc actcgcacca ggtcgaggac ctgcacgagg
        69551   aactgctgga ggtgctggcg gagctggcgc cgcgcacgtc ggaggtgccg
        69601   ttcttctcga ccgtgaccgg cgactggctg gacaccgcgc ggatggacgc
        69651   cggctactgg ttccgcaacc tgcgcggacg ggtgcggttc gcggacgcgg
    25  69701   tggcggacct gctggcggcg gagtaccgcg cgttcgtcga ggtcagctcg
        69751   cacccggtgc tgacgatggc ggtcttggac ctgatcgagg aggccggggt
        69801   cacggccgtc gcgaccggca ccctgcgccg tgaccagggt ggcgcgggcc
        69851   gcttcctgct gtcggccgcc gaggtcttcg tgcgcggtgt ggacgtggac
        69901   tgggcggggg cgttcgaggg gaccggtgcg gcccgggtcg acctgcccac
    30  69951   ctacgccttc cagcgcgagc ggtactggaa cacccgcacc gccgccgacc
        70001   gcaccccggc cgacgccccg atggacgccg aattctgggc cgccgtcgaa
        70051   caggcggacg tctccgcgct gaccgccgcg ctcggcaccg acgaggactc
        70101   cgtcgccgcc atcctgcccg gcctcacctc ctggcgccgg gcccgctccc
        70151   agcgcaccac cctcgactcc tggcgctacc gcgtcacctg gacgcccctc
    35  70201   gcccaggtgc cccgcgccac cctgaccggc acctggctgc tggtcaccac
        70251   cgacggcatc gacgacaccg atgtggcagg ggcgttggag agctacggcg
        70301   ccgaggtgcg ccggctggtc ctggacgagg agtgcaccga ccgcgccgtc
```

```
        70351  ctgcgggagc ggctggccgg cgcggaggac gtgaccggca tcgtctccgt
        70401  cctcgccgcc gccgaggacg acgccgcacg ccaccccggc ctcacccggg
        70451  gactcgcgct caccgtctcc ctcgtccagg ccctgggcga cgccgaggcg
        70501  accgcgccgc tgtggttcct gacccgcggc gccttcgcca ccggcccgtc
   5    70551  cgaccccgtc acccggcccc tgcagagcca gatcgcgggc gtcggctgga
        70601  ccaccgcgct ggagcacccg cagcgctggg gcggcaccgt ggacctgccc
        70651  gacaccctcg acgcccgggc cgcccagcgg ctcgccgccg cgctgtccgg
        70701  cgccctcggc gccgaggacc agctcgccgt ccgcgccgcc ggggtactgg
        70751  cccgccgcat cgtgcgtgcc ggacaccgcg ccggacgacc ggcacggacc
   10   70801  tgggcgccgc gcggcaccac cctgatcacc ggcggctccg gcaccctcgc
        70851  cccgcagctc gcccgctggc tggccgaacg cggcgccgag cacgtggtgc
        70901  tggtcagccg gcgcggtgcc gacgcccccg gagcgcccga actcatcgcg
        70951  gaggcagccg agtcgggcac cgaggtgacc gtcgccgcct gcgacatcac
        71001  cgaccgcgac gcggtcgccg cgctgctggc cgacctcacg gccgacggcc
   15   71051  gcaccctgcg caccgtcatc cacgccgccg ccgccatcga gctgtccgcg
        71101  ctcgccgaca ccaccgtggc ggagttcgcc gacgtcgtgc acgccaaggt
        71151  caccggcgca cggatcctcg acgaactgct cgacgacgcg gaactggacg
        71201  acttcgtcct gtactcctcc accgccggca tgtggggcag cggcgtgcac
        71251  gccgcctacg tcgccggcaa cgcctatctg tccgcgctcg ccgagcagcg
   20   71301  ccgcgcccgc ggactgcgca ccacctccat ccactggggc aagtggcccg
        71351  acgaccgggc acgcgagctg gccgacccgc accggatccg ccgcagcggt
        71401  ctggagtacc tcgaccccga gctggcgctc accgcgctcc agcacgtcct
        71451  ggacgacgac gagaccgtca tcggcctcat ggacatcgac tgggacacct
        71501  accacgacgt gttcaccgcg ggccggcccg cgcacctctt cgaccagatc
   25   71551  cccgaggtgc ggcgccgcct cgaccaggca tccgtcccgg accccgcggg
        71601  cccggccgcc gacggcctgg ccgcccgcct gcacggcctc gccgccgccg
        71651  aacaggaccg gctgctgctc accctggtcc gcaccgaggc cgccgccgtc
        71701  ctcggccacg cctcggccga gtccttcccc gagcgccgcg ccttccgtga
        71751  cctcggcttc gactcggtca ccgccgtgga cctgcgcaac cggctcgtgg
   30   71801  ccggcaccgg actgcggctg ccctcgacga tggtcttcga ccaccccaac
        71851  tgcgcggcgc tcgccgcgtt cctgaagacg acggcgctcg gcgtccccgg
        71901  cgccgcaccg cagcagcacg ccgctaccgg caccccggcc gacgacgacc
        71951  cgatcgccgt gatcggcatg agctgccgct accccggcgg cgccgccacc
        72001  cccgaggaac tgctgcggct cgccctcgac ggcgccgacg tcatctcgga
   35   72051  gttccccgcg gaccgcggct gggacgcccg gggcctgtac gacccggacc
        72101  ccgaccgccc cggccacacc tactccgtcc agggcggctt cctccacgag
        72151  gccgccggct tcgatcccgg cttcttcggg atctccccgc gcgaggcggt
```

```
        72201  cgccatggac  ccgcagcagc  ggctcctgct  ggagacctcc  tgggaggcgt
        72251  tcgaacgcgc  cggtatcgac  cccgcgtcac  tgcgcggcag  cgccgccggc
        72301  accttcttcg  gcgccagcta  ccaggactac  tcctccaccg  tgcagaacgg
        72351  cacgggggag  tccgaggcgc  acatggtgac  cggcaccgcg  gccagtgtcc
  5     72401  tgtccggccg  ggtctcctac  ctgctcggcc  tggagggccc  cgcggtcacc
        72451  gtggacaccg  cctgctcctc  ctcactggtc  gccctgcacc  tggcctgcca
        72501  gtccctgcgc  gacggcgaga  gctccctcgc  gctggccggc  ggtgcggccg
        72551  tgatggccac  cccgcacgcg  ttcgtcggct  tcagccggca  gcgtgccctg
        72601  gccaaggacg  gccgctgcaa  gccgttctcc  gacaccgccg  acggcatgac
  10    72651  gctcgccgag  ggcgtcggcg  tcgtcctgct  ggagcgcctg  tcccacgccc
        72701  gcgccaacgg  gcaccgggta  ctggccgtga  tccgcggttc  cgccgtcaac
        72751  caggacggcg  cctccaacgg  cctgaccgcg  cccaacggcc  cgtcccagca
        72801  gcgcgtcatc  cgccaggcgc  tcgccaacgc  cggcctgacc  ggcgccgacg
        72851  tcgacgcggt  cgaggcgcac  ggcaccggca  ccaagctggg  cgaccccatc
  15    72901  gaggcccagg  ccctgctcgc  cacctacggc  caggaccgcg  acgccgaacg
        72951  gccgctgctg  ctgggctcgg  tgaagtccaa  catcggccac  acccaggccg
        73001  cggccggcgt  cgccggcgtc  atcaagatgg  tgctggccat  ggacgccggc
        73051  gaactgcccg  gcaccctgca  cctcgacgcg  ccctccagcc  acgtcgactg
        73101  gaccgccggc  gccgtggaac  tgctgcgcgg  gcgcaccccg  tggcccgaga
  20    73151  gcgggcgccc  ccgccgggcc  ggtgtctcct  cgttcggcat  cagcggcacc
        73201  aacgcccacc  tgatcctcga  acaggccccg  gccaccgagc  cgccagccga
        73251  ccccgaccgc  ctccgggaca  ccgccaccga  caccgtcgtc  ccctggccgc
        73301  tcgccgccaa  gtccccggcc  gccctgcgcg  cccaggccgc  ccggctcctc
        73351  gccaccgtcg  agcacgaccc  cgacctcccg  cccgccccg  tgggccacgc
  25    73401  cctggccacc  acccgcgccg  ccctcgaaca  ccgcgccgtc  gtcgtcggcg
        73451  agcgccgcga  ggacttcctg  cgcggcctgg  ccgccctgtc  caccggcgcc
        73501  tcgacgccg  gcctggtcag  cggcatcgcc  ggccccgacc  ccgagggagc
        73551  ggtcttcgtc  ttccccggcc  agggatccca  gtggtgggga  atgggccgcg
        73601  aactcctcgc  cacgtccgag  gtgttccgca  ccgcgatcga  tgactgcgcg
  30    73651  acggccctcg  ccccgtacgt  cgactggtcg  ctgcacgacg  tcctggccgg
        73701  cgagggcgac  cccgccctgc  tggagcgggt  ggacgtggtc  cagcccgcgc
        73751  tgttcgccat  gatggtcggg  ctgtccgcgc  tctggcgctc  ccacggcgtc
        73801  gtcccggcgg  ccgtggtcgg  ccactcgcag  ggcgagatcg  ccgcggcctg
        73851  cgtcgccgga  gccctcagcc  tggccgacgc  cgcccgcgtg  gtggcgctgc
  35    73901  gcagccaggc  actgccgcaa  ctgtccggac  gcggcggcat  gatgtcggtc
        73951  tccgcccccg  tagagcgggt  caccgcactc  ctcgccccgt  ggcaggaggc
        74001  gctgtccgtc  gccgcggtca  acggcccctc  gtccgtggtc  gtctccggcg
```

```
      74051  acaccgacgc gctcgacgcc ctgcacaccg cctgccagga acagggcgtg
      74101  cgggcccgca aggtgtccgt ggactacgcc tcgcacgggc ggcacgtcga
      74151  ggccgtccgc gacgaactcg cccgcgtcct cgcgccggtc gacccgcgcg
      74201  cccccgaggt gccgttctac tcgacggtca ccggcgaccg cgtggacgac
   5  74251  gccgccttcg acggcgccta ctggtacacc aacctccgcc agaccgtccg
      74301  catggaggag gccacccgcg ccctcctcgc cgccggacac cgcgtcttca
      74351  tcgaggtcag cccgcacccg gtgctcgccg ccccgatcca ggagacgcag
      74401  gaggccgtag cggaggccac cggcgggtcc gcggtggtcc tcggctcgct
      74451  ccgccgcgac gagggcggcc cgcggcgctt cctgacgtcg ctcgccgagg
  10  74501  cccacaccca cggcgccccg gtcgactgga ccaccacctt cgcccggtcc
      74551  gcctaccagc cggtggacct gccgacctac cccttccaac gacaggactt
      74601  ctggcccgag gcccggcccg ccaccccggc cgccggcgcc gacgcgtccg
      74651  acgccgcgtt ctggcaactg gtcgagaacc aggacctcgc cgcgctcgcc
      74701  gacgcgctcg gcgtccccgc cgacgacgag cacaccgcgc tcggcaccgt
  15  74751  gctgccggcc ctgtccgcct ggcgcgccaa ggcccaggcc cgcacccgga
      74801  tcgacgaact ccgctaccac gtccagtgga cccgggtcgc cgagcccgcg
      74851  gcggcccccca ccaccggccg gctgctggtc gccgtcccgc cggaccacgc
      74901  cgacgccccc tgggtcgccg cggcgctcga cgccctgggc accgacaccg
      74951  tccgcttcga ggccaagggc accgaccgcg cgggatgggc cgcacagatc
  20  75001  gcccaactcg tcgaggacgg cgaggagttc accggcgtgg tgtcgctgct
      75051  ggccgccgcc gaggatctcc acccggactt cggctcggta ccgctggggc
      75101  tggggcagac cctcgtcctc gtccaggccc tcggcgacgc cggcctgacc
      75151  gcgcccctgt ggtgcctgac ccgcggcgcc gtcgccaccg gccgcgacga
      75201  cgccctcgac agcccgaccc agggcgccct gtggggcctc ggccgggtcg
  25  75251  tggccctgga acaccccgac cgctggggcg gcctgatcga cctgcccgcc
      75301  accctcgacg cccgcgccgc ggcccgcctc accggcctgc tcgccgaccc
      75351  cgccggtgag gaccaactcg ccgtccgcgc caccggcgtg ctcgcccgcc
      75401  gcatggtgca cgccgcgccg tccgcgcccc gcaccgggcg ccgctggcgg
      75451  ggccgcggca cctgcctgat caccggcggc accggcggca tcggcggccg
  30  75501  ggtcgcccgc tggatggccg agcacggcgc cgcccacctg gtcctgacca
      75551  gccggcgcgg cccggacgca cccggcgccg ccgcactccg ggccgaactg
      75601  gaggccctgg gcgccgcgt caccctcgcc gcctgcgacg tcgccgaccg
      75651  cgacgccctg ccgccctgc tggccgacct ccccgccgac cagccgctca
      75701  cctccgtctt ccactccgcc ggcgtggccg acggggacgc ccgggcagcc
  35  75751  gacctgaccc tagatcagct cgacgcgctg ctgcgcgcca aactgaccgc
      75801  cgcccaccac ctgcacgagc tgaccgcccc cctcgacctc gacgcgttcg
      75851  tgctcttctc ctccggcgcc gcggtctggg gcagcggcgg ccagcccggc
```

```
      75901  tacgccgccg  ccaacgccta  cctcgacgcc  ctcgccgccc  accgcaggtc
      75951  cctcgacctg  cccggcgcgt  ccgtcgcctg  gggcacctgg  ggcgaggtcg
      76001  gcatggccac  cgtccccgag  gtccacgagc  gactgcaccg  ccaaggggtc
      76051  cgcgccatgg  aaccggacca  cgcgatcggc  gcgctccagc  agatgctgga
 5    76101  ggacgacgac  accaccctcg  ccgtgaccct  catggactgg  gaggcgttcg
      76151  caccgagctt  caccgcgacc  cgacccagcg  ccctgttcag  cacggtgccc
      76201  gaagccgtcc  gcgcggtgac  cggcgacccg  ggcaccacgg  ccggcgacga
      76251  cgtggactcc  gcgaccccgc  cgctccgccg  ccacctggag  gagctgtccg
      76301  ccgccgagcg  cggccgggcc  ctggtcgagg  cggtccgcgc  cgaggcgtcc
 10   76351  gcgaccctcg  gccacgacac  ccccgacgcc  atccccgccg  gccgtgcctt
      76401  ccgcgacgtc  ggcttcgact  cggtcaccgc  cgtcgaactg  cgcaaccggc
      76451  tgcgcaccgc  cctcggcctg  ccgctgccgg  ccgcgctcgt  cttcgaccac
      76501  cccacccccca  cggccctcgc  cggccacctc  ggcgcgctgc  tcttcggcac
      76551  cgcccccgag  gacgccggca  ccggccgccc  cgacgacccc  gacgcccgca
 15   76601  tccgcgaggc  gctcgccacc  gtccccatcg  gacggctgcg  caaggcgggc
      76651  ctcctcgaca  tggtgctgaa  actcgccgac  ggagacgcga  ccgacgcccc
      76701  cgcccccgag  gccgacgccc  cctcggaatc  cctcgacgac  atggacgccg
      76751  aagccctgct  gcggctggcc  accgagaact  cggcgaactg  aaagagagct
      76801  ggagcccacc  atgagcacga  accccgacaa  gtacgtcgag  gcactccggt
 20   76851  cgtccctgaa  ggagatcgag  cggctgcgcc  ggcagaacga  acagctggtg
      76901  gcggcggcgg  tcgagcccgt  cgcggtcgtc  ggcatcggct  gccgcttccc
      76951  cggcggcgtc  acctccccccg  aggacctgtg  ggagctggtc  gccgaggggc
      77001  gcgacgtcat  cgggccgttc  ccgcaggacc  gcggctggga  cctggagaag
      77051  ttggccggcg  gcggcgaggg  cggcagcctc  gcgcaggtcg  gcggcttcgt
 25   77101  cgaggacgcc  gccggcttcg  accccggctt  cttcggcatc  tccccgcgcg
      77151  aagcggtcgc  catggacccg  cagcagcgca  tcctgctgga  gatcacctgg
      77201  gaagccctgg  aacgcgccgg  catcgacccg  tccaccctgc  gcggtacccc
      77251  caccggcgtc  ttcgtcggca  ccaccggcca  ggactacggc  gaggtcatca
      77301  aggcgtccgc  cgaggacgtc  gaggtctact  cgaccaccgg  ccacgccgcc
 30   77351  agcgtcatct  ccggccggct  ctcctacacc  ctcggcgccg  agggcccggc
      77401  cgtcaccgtc  gacaccggct  gctcctcgtc  cctggtcgcc  ctgcactggg
      77451  ccgtccaggc  gctgcgcggc  ggcgagtgct  ccatggcccct  ggccggcggc
      77501  gcgtccatca  tggccacccc  gggcccgttc  gtcgccttca  ccgcgcagag
      77551  cggcctggcc  gccgacggcc  gctgcaagcc  cttctccgac  cgggccgacg
 35   77601  gcaccggctg  gggcgagggc  gccggcatgc  tggtcctgat  gcggctctcc
      77651  gacgcccagc  gcgagggccg  cccggtcctc  gccgtgctgc  gcggctccgc
      77701  catcaaccag  gacggcgcct  ccaacggcct  gaccgctccc  aacggcccct
```

```
        77751  cccagcagcg cgtcatccgc gccgcgctgg acagcgccca cctcaccgcc
        77801  gccgacatcg acgccgtcga ggcccacggc accggcacca ccctcggcga
        77851  cccgatcgag gcccaggcgc tcctggcgac ctacggacag gaccggccgc
        77901  ggcccctgtg gctcggctcg gtgaagtcca acatcggcca cacccaggcc
   5    77951  gcctccggtg ccgccggcgt gatcaaaatg atcatggcgt tgcagcgcgg
        78001  cgtgctgccg cgcagcctgc acgccaccga acccaccacg gacgtcgact
        78051  ggaccgccgg ctccgtcgac ctcctcgacg agacggtcgc ctggcccgag
        78101  accggacgcg cccgccgcgc cggcgtctcc tccttcggca tcagcggcac
        78151  caacgcccac gtcatcctcg aacaggcccc caccgccccc gaagagccca
   10   78201  ccaccgaacc caccgtccgc cccgccgtcg tcccgtgggc gctctccgcc
        78251  cgcaccgccg ccgccctcga cgcccagcgc gcccgcctca ccggccacct
        78301  cgccgacacc cccgacgccg accccctcga cgtcggctac gcgctcgccg
        78351  acggacgcgc caccttcgaa caccgcgccg tcctgctccc cgacggcacc
        78401  gaactcgccc acggaaccgc cggcgaaggc ccctgcgccg tcctcttctc
   15   78451  cggccagggc tcccagcgcc cgggcatggg acgcgaactc cacgcccgct
        78501  tcccggtgtt cgccgcggcc ttcgacgaga tcacagcgct cctcgacacc
        78551  cacctcgacc gcccgctgcg cgaggtcgtc tggggcaccg acgccgacct
        78601  gctgaacgac accggctggg cccaacccgc cctgttcgcc gtcgaggtcg
        78651  ccctctaccg cctggtcgcg tccctcggcg tcacccccga cttcgtcggc
   20   78701  ggccactcca tcggcgagct cgccgccgcg cacgtcgccg gggtcctctc
        78751  cctcgaagac gcctgcaccc tcgtcgccgc ccgcgcccgc ctcatgcagg
        78801  ccctgccgcg cggcggcgcg atgctcgcga tccgcgccac cgaggacgag
        78851  gtcacccccc acctcaccga cgacgtctcg atcgccgccg tcaacgggcc
        78901  cacctccgtc gtcgtcgccg gcaccgagga agccgtcgcc gcgatcgggg
   25   78951  cgcgcttcac cgcccaggac cgcaagacca cccggctgcg ggtcagccac
        79001  gccttccact cgccgctcat ggacccgatg ctggcggaat tccgcgccgt
        79051  cgccgcgggc ctgacctacc acgagccgcg catcccggtc ctctccaacc
        79101  tcaccggcac cgtcgccgcc gtcgccgacc tgtgctccgc cgactactgg
        79151  gtccgccacg tccgcgaggc ggtccgcttc gccgacggcg tcaccgccct
   30   79201  caccgaccgc ggcgtgacca cgctcgtcga actcggcccg gacggcgtgc
        79251  tgtccgccat ggcccaggaa tccctgccgg acggcgccgc cgccgtgccg
        79301  ctgctgcgca aggaccgccc cgaggagctc tccgccgtca ccggcctggc
        79351  ccgcgcccac gtccgcggcg tcacggtccg ctgggccggc ctcttcgacg
        79401  gcaccggcgc gcgccgcgcc gacctgccca cctaccccct ccagcaccag
   35   79451  cggttctggc cgaccgcggc ccgcgccgcc caggacgtca ccgccgcggg
        79501  actgggcgcc gccgaccacc cgctgctcgg cgccaccgtc gaactcgccg
        79551  acggggccgg ctacttgttc accagccggc tctccgtccg gacccacccc
```

```
        79601  tggctcgccg accacggggt ccagggccgg gccctgctgc ccggcaccgc
        79651  cttcgtcgaa ctggccgtcc gcgccggcga cgaggccggc tgcgaccgcg
        79701  tcgaggaact gaccctggcc gcccccctgg tgctgcccga gcgcggcggc
        79751  gtccaactcc aggtccgcgt cggcgccccc gacgccgccg gccgccgcac
    5   79801  cctcggcatc ttctcccgcg tcgaggacgg cttcgacctg ccctggtcgc
        79851  aacacgccac cggcgtcctg accgccggcg ccggcgcccc cgaccccacc
        79901  ttcgacgcca ccgtctggcc ccccagcggc gccgaacccg tcgacctcac
        79951  cggcgcgtac gagcgcctgg ccgcactcgg cttccagtac ggccccgcct
        80001  tccagggcct gcgcgccgcc tggcgccgcg acaccgaggt ctacgccgaa
    10  80051  gtggccctgc ccgacggcgc ggacaccgac cccgccgcct tcggactgca
        80101  cccggccctg ctggacgccg cacaacacgc cgccgcctac gccgacctcg
        80151  gcgccatcag ccgcggcggc ctgccgttcg cctgggaagg cgtctcgctc
        80201  gccgccgccg gcgccaccac cgtccgcgcc cggatcgccc cggccggcga
        80251  ggacaccgtc accatcgccg tctacgacgc cgccggcggc accgtgctgt
    15  80301  ccgtcgactc cctggtctcc cgcgaggtcc ccgccgacgc acccggcgcc
        80351  gccggcaccg tccaccgcga ctccctcttc cacgtcgagt ggaccccgct
        80401  ccagggccgc ccgggccccg caccggccac cgtcgccgtc ctcggccccg
        80451  acccggacgc cctcgccgac accctccgcg ccaccggcat ccggaccacc
        80501  gccccccgcg acctggccgc cctcgccgac gccgaagggc ccgtccccga
    20  80551  cctggtcgtc accaccctca ccaccacccc gggcgccccc gtccccgacg
        80601  ccgcgcacgc caccaccgcc gccgtcctcg ccctcgccca acagtggctc
        80651  gccgacgacc gcttcgccga cgcccgcctg gtcctcgtca cccgcggcgc
        80701  caccgacggc accgaccccg ccgccgcggc cgccggcggc ctgatccgca
        80751  ccgcccgcac cgagaacccc ggccgtttcg ccctcctcga cctcgccccc
    25  80801  gacaccggcc ggcccgaccc cgagaccctg gccaccgccc tggccgccag
        80851  ccacgacgag cccgacctcg ccgtccgcgg caccgacgtg cacgccgccc
        80901  gcctggcccg tgtcccgctc gccaccgaac ccaccacctg gaacccggac
        80951  ggcaccgtcc tgatcaccgg cggcaccggc ggcctgggcg cggtcctcgc
        81001  ccgccacctg gtcgccaccc acggcgtccg ccacctgctg ctcgccagcc
    30  81051  gccgcggccc ggccgccgac ggcgccgacg acctgacggc cgaactcacc
        81101  gggctcggcg ccaccgtcca catcgccgcc tgcgacgtcg ccgacccggc
        81151  cgccctcgcc gacctgctcg gcaccgtccc ggccgggcac ccgctcaccg
        81201  tcgtcgtcca caccgccggc gtcgtcgacg acggcgtcct cggctccctc
        81251  accccgcagc gcctggacac cgtcctgcgg cccaaggccg acgccgcctg
    35  81301  gcacctgcac gaggcgaccc gccacctcga cctggacgcc ttcgtcctct
        81351  tctcgtccgt cgccgccacc ctcggcagcc ccggacaggc caactacgcc
        81401  gccggcaacg ccttcctgga cgccctcgcc gcccggcgcg ccgccaccgg
```

```
81451    cctgcccgcc  acctccctcg  cctggggccc  gtggacccag  agcgtcggca
81501    tgacaagcag  cctgtccgac  ctcgacgtcg  agcgcatcgc  ccgctccggc
81551    atgcccccgc  tgaccctgga  acagggcacc  gccctcttcg  acgcggccct
81601    ggccgccggg  cccgccgccc  tcgccccggt  ccgcctcgac  ctgcccgtcc
81651    tgcgcaccca  gggcgacatc  gccccgctgc  tgcgcggcct  gatccgcacc
81701    cccgtgcggc  gcaccgccgc  ccaggtctcg  cagaccgccg  acggcctcgc
81751    ccagcggctc  gccggcctcg  acgccgccgc  ccgccgggaa  gccctcctgg
81801    aactcgtccg  cacccagatc  gcccaggtcc  tcggccacgc  ggacgccacc
81851    gaggtggaga  ccggccgcca  gttccaggac  ctcggcttcg  actccctcac
81901    cgccgtcgaa  ctccgcaacg  ccctgaacac  cgccaccggc  ctgcggctgc
81951    ccgccaccat  ggtgttcgac  tacccgacac  cacacgccct  cgccgaccac
82001    ctgcgcgacg  aactcctggg  caccgaggcc  gagtcgacca  ccgccgtccc
82051    cgtgccgacc  cgtaccgccg  gcaccgacga  cccgatcgtc  atcgtcggca
82101    tggcctgccg  ctaccccggc  ggcatcgcct  cacccgagga  cctctggcgc
82151    ctggtcagcc  agggcgccga  cgccactggc  ccgttcccca  ccaaccgcgg
82201    ctgggacctg  gacaacctct  acgaccccga  ccccgaccgc  ccgggccgca
82251    cccacgtccg  cgccggcggc  ttcctgcacg  acgccggctc  cttcgacgcc
82301    gacttcttcg  ggatgagccc  gcgcgaggcg  atggccaccg  actcccagca
82351    gcgcctgctg  ctcgaactct  cctgggaagc  cgtcgaacgc  gccggcatcg
82401    accccgcctc  actgcgcgac  tccggcaccg  gcgtcttcgc  cggcgtcatg
82451    tacaacgact  acggcaccac  cctgaccggc  gacgagtacg  aggcgttccg
82501    cggcaacggc  agcgccccga  gcgtcgcctc  cggccgcgtc  tcctacaccc
82551    tcggcctgga  aggcccggcc  gtcacggtgg  acaccgcctg  ctcttcctct
82601    ctggtcgccc  tgcactgggc  ggcgcaggcg  ttgcgggcgg  gggagtgctc
82651    gttggcgttg  gccggtggtg  tgacggtgat  gtcgacgccg  agcacgttcg
82701    tggagttctc  gcggcagcgg  ggtctggcgc  ctgatggtcg  ttcgaaggcg
82751    ttcgccgagg  ccgcggacgg  cgtggcctgg  tccgagggcg  tcggcatgct
82801    ggtcctggag  cggcagtcgg  acgcggtgcg  caacggtcac  gagatcctgg
82851    ccgtggtgcg  cggctcggcg  gtcaaccagg  acggtgcgtc  caacggtctg
82901    accgcgccca  acgcccgtc   ccagcagcgg  gtgatccgtc  aggcgttggc
82951    cagtggcggc  ctgtccacgg  ccgacgtgga  cgccgttgag  gcgcacggca
83001    cgggtacgac  gctcggtgac  ccgatcgagg  cccaggcgct  cctggccacc
83051    tacggtcgcg  accgcgaccc  cgagaacccg  ctgctgctcg  gctcgatcaa
83101    gtccaacatc  ggtcacaccc  aggcagcggc  cggtgtcgcc  ggtgtcatca
83151    agatggtcat  ggcgatgcgg  cacggcgtgc  tgccgcagac  cctgcatgtc
83201    gacgcgccgt  cctcgcacgt  cgattggagc  gtcggcgccg  tcgaactgct
83251    caccgagcag  accgcctggc  cggagaccgg  ccgggcccgt  cgcgccggtg
```

```
83301  tctcctcctt cggcatcagc ggcaccaacg cccacgtcgt catcgagcag
83351  tccccgaccg ccgtccccgc cacgcccgcg tccgccgacc ggtccgtcga
83401  ggaaccgccg gccgtcccct gggccctgtc cggcaagacc cccgacgccc
83451  tccgcgacca ggccgcccgc ctcctcgccc acgtcgaggc ccaccccgca
83501  ctgcgccccg tcgacatcag ctactccctg atcgccaccc gcaccgcctt
83551  cgaccaccgc gccgtcgtcc tcggcaccga ccgcgccgag gccctgcgcg
83601  ccctcaccgc cctcgccgcc ggcgagaccg acccggccgc cctcaccggc
83651  accgtccgca ccggccgcac cgccttcctc ttctccggcc agggctccca
83701  acggctcggc atgggccgcg tcctctacga gcggttcccc gccttcgccg
83751  aagccctcga caccgtcctc accgccctcg acgcggaact cggccacccc
83801  ctccgcgaca tcatctgggg cgaggacgct caactcgtcg accggaccgg
83851  ctacacccaa cccgccctgt tcgccatcga ggtggcactc ttccgcctcc
83901  tggaagcctg gggcatcaca ccggacttcg tggccggcca ctccatcggc
83951  gagatcgccg ccgcacacgt cgccggcgtg ctctcccctcg gcgacgcctg
84001  ccgcctcgtc gtggcccgcg ccgtgctgat gcagtcgctg cccgaaggcg
84051  gcgcgatgat cgccgtccag gccaccgagg acgaggtcct gcccctcctc
84101  accgacgacg tctcgatcgc cgccgtcaac agcccgacct ccgtcgtcgt
84151  ctccggctac gagaacgcca ccctcgccgt cgcccggcac ttcgccgacc
84201  agggccgccg caccacgcgg ctgcgcgtca gccacgcctt ccactcgccg
84251  ctgatggcgc cgatgctcga cgacttccgc gccgtcgtcg agagcctcac
84301  cttcaccgcc cccacgaccc ccgtcgtctc caacctgacc ggcgaactgg
84351  ccccggccga ggcgctctgc tcggccgact actgggtccg gcacgtccgc
84401  gaggcggtcc gcttcgccga cggcatccgc accctcgccg accgcggcgt
84451  caccaccttc gtcgaactcg ccccgacag cgtgctgtcc gccatggccc
84501  aggagtccgc ccccgaaggc gccggcacca tcccgctcct cgccgcgcgac
84551  cggcccgagg aacaggccgt cctggccgcc ctctgccacc tccaggtgct
84601  cggcgtcgag gccgactggt ccgccacctt ccgcggcctc gaccccgtcc
84651  gcgtcgacct gccgacctac gccttccagc accgctggtt ctggcccgcc
84701  gcccgacccg cccgccccga cgacgtccgc gccgccggcc tgggcgccgc
84751  cgaacacccc ctcctcggcg ccgccgtgca actccccgac gacgacggcg
84801  cactcttcac cggccgcctc tccctgcgca cccacccgtg gctggccgac
84851  cacaccgtcc tgggcaccgt cctgctcccg ggcaccgcac tggtggaact
84901  cgccgtccgc gcgggcgacg agaccggcag cggccacctc gaagaactca
84951  ccctcgccgc gccctgacc ctccccgagg acggcgccac cctcctccag
85001  gtccgcgtcg gatccgccga cgacaccggc cgccgcaccg tcaccgtcca
85051  cgcccgcccc gacgacaccg ccgaccgcac ctggacgctg cacgccaccg
85101  gtgtgctcgc caccacgcca ccggccgccg cggcgttcga caccacggtc
```

	85151	tggccgcccg	ccgacgccga	accccctcacc	accgacgact	gctacgcaca
	85201	cttcaccacc	caccgcttcg	cctacggccc	cgccttccag	ggcctgcggg
	85251	ccgcctggcg	cgccggcgac	gtgctgtacg	ccgaggtcgc	cctgccggag
	85301	tccgccaccg	acgaagcggc	cgccttcggc	ctgcacccgg	cgctcctgga
5	85351	cgccggcctg	cacgccgcgc	tcctcgccga	cgaccgcgac	accggactcc
	85401	cgttctcctg	ggaaggcgtc	actctgcacg	cctccggcgc	caccgcgcta
	85451	cgcgtccggc	tcgccccgaa	cggccccaac	ggcctgtccg	tcaccgccgc
	85501	cgaccggcc	ggcaacccg	tcgccaccgt	cacccgcctg	ctcgcccgcc
	85551	ccctggacgc	cgagcagttg	accatccaca	gcgccctgac	ccgcgacgcg
10	85601	ctcttccacc	tggactggac	cccggtcccg	cttcccgaca	ccgccaactc
	85651	cgcgccgccg	gccctcctcg	gcccggacac	cgccgtgctc	gccgacgccc
	85701	tcggcgaccc	ggccgtcgca	cgccacgcaa	ccctcgacga	cctcctggcc
	85751	ggggacacca	ccccgcccgc	cacggtcctc	gtccccctcg	gcgccccact
	85801	cgacggcgac	accgcgcagc	acgcgcacgc	cctcacccgc	agcgcgctga
15	85851	ccctcgtcca	gcagtggctc	gccaccgacc	gcctcgccga	ctcccgcctg
	85901	gtcttcgtca	cccacggagc	cgtcgccacc	gacgacgcgc	cccccaccga
	85951	cctggccgcc	gccgcggtct	ggggcctgat	ccgctccgcg	cagaccgaga
	86001	accccggcac	cttcacccctc	ctcgacctcg	acaccgagcc	cgactcgacc
	86051	accgcgctca	gccgcgccct	gaccctcgac	gaaccacagc	tcctcctccg
20	86101	cgccggccgc	gcccgcgccg	cccgcctcac	ccgcacccccc	gcccccacca
	86151	ccaccaccca	cacgccgtgg	tccgcggacg	gaacggtgtt	ggtgacgggt
	86201	ggtacgggtg	gtctgggtgg	gttggtggcc	cggcatctgg	tgcggtcgtg
	86251	tggggtgcgg	catttgttgt	tgaccagtcg	ttctggtgtg	ggtgctgcgg
	86301	gtgcggccgg	gttggtcgcg	gagttggagt	cgttgggcgc	gcgggttgtg
25	86351	gttgcggcgt	gtgatgtggg	tgatggctcg	gctgttgcgg	agttggttgc
	86401	cggtgtgtcg	gagtcgtatc	cgttgtctgc	ggtggtgcat	gcggctggtg
	86451	tgttggatga	cggtgtggtg	ggttcgttga	cgccggagcg	gttggctgcg
	86501	gtgttgcgtc	cgaaggtgga	tggtgcgtgg	aacctgcatg	aggcgacgcg
	86551	tggtctggat	ctggacgcgt	ttgttgtctt	ctcgtctgtt	gcgggtgtgt
30	86601	tcgggggtgc	gggtcaggcc	aactatgcgg	cgggtaatgc	gttttggac
	86651	gcgttgatgg	ttcatcgggt	ggctggtggg	ttgcctggtg	tgtcgttggc
	86701	gtggggtgct	tgggatcagg	gtgtggggat	gacggcgggg	ctgacggagc
	86751	gggatgtccg	tcgtgctgct	gagtcgggta	tgccgttgtt	gacggttgat
	86801	cagggtgtgg	cgttgttcga	tgcggcgttg	gcgacgggga	gtgccgcgtt
35	86851	ggtgccggtc	cgtctggacc	tggccgcact	gcgcacccgg	ggcgacatcg
	86901	caccgctcct	ccgcggcctc	gtccgcgcac	cgctgcgccg	caccgcggcc
	86951	accggcctcg	ccaccggcgc	ggacaccggc	ctcgtccaac	ggctcggccg

```
      85151  tggccgcccg ccgacgccga acccctcacc accgacgact gctacgcaca
      85201  cttcaccacc caccgcttcg cctacggccc cgccttccag ggcctgcggg
      85251  ccgcctggcg cgccggcgac gtgctgtacg ccgaggtcgc cctgccggag
      85301  tccgccaccg acgaagcggc cgccttcggc ctgcacccgg cgctcctgga
   5  85351  cgccggcctg cacgccgcgc tcctcgccga cgaccgcgac accggactcc
      85401  cgttctcctg ggaaggcgtc actctgcacg cctccggcgc caccgcgcta
      85451  cgcgtccggc tcgccccgaa cggccccaac ggcctgtccg tcaccgccgc
      85501  cgacccggcc ggcaaccccg tcgccaccgt cacccgcctg ctcgcccgcc
      85551  ccctggacgc cgagcagttg accatccaca gcgccctgac ccgcgacgcg
  10  85601  ctcttccacc tggactggac cccggtcccg cttcccgaca ccgccaactc
      85651  cgcgccgccg gccctcctcg gcccggacac cgccgtgctc gccgacgccc
      85701  tcggcgaccc ggccgtcgca cgccacgcaa ccctcgacga cctcctggcc
      85751  ggggacacca ccccgcccgc cacggtcctc gtccccctcg gcgccccact
      85801  cgacggcgac accgcgcagc acgcgcacgc cctcacccgc agcgcgctga
  15  85851  ccctcgtcca gcagtggctc gccaccgacc gcctcgccga ctcccgcctg
      85901  gtcttcgtca cccacggagc cgtcgccacc gacgacgcgc ccccaccga
      85951  cctggccgcc gccgcggtct ggggcctgat ccgctccgcg cagaccgaga
      86001  accccggcac cttcaccctc ctcgacctcg acaccgagcc cgactcgacc
      86051  accgcgctca gccgcgccct gaccctcgac gaaccacagc tcctcctccg
  20  86101  cgccggccgc gcccgcgccg cccgcctcac ccgcaccccc gcccccacca
      86151  ccaccaccca cacgccgtgg tccgcggacg gaacggtgtt ggtgacgggt
      86201  ggtacgggtg gtctgggtgg gttggtggcc cggcatctgg tgcggtcgtg
      86251  tggggtgcgg catttgttgt tgaccagtcg ttctggtgtg ggtgctgcgg
      86301  gtgcggccgg gttggtcgcg gagttggagt cgttgggcgc gcgggttgtg
  25  86351  gttgcggcgt gtgatgtggg tgatggctcg gctgttgcgg agttggttgc
      86401  cggtgtgtcg gagtcgtatc cgttgtctgc ggtggtgcat gcggctggtg
      86451  tgttggatga cggtgtggtg ggttcgttga cgccggagcg gttggctgcg
      86501  gtgttgcgtc cgaaggtgga tggtgcgtgg aacctgcatg aggcgacgcg
      86551  tggtctggat ctggacgcgt ttgttgtctt ctcgtctgtt gcgggtgtgt
  30  86601  tcggggggtgc gggtcaggcc aactatgcgg cgggtaatgc gttttttggac
      86651  gcgttgatgg ttcatcgggt ggctggtggg ttgcctggtg tgtcgttggc
      86701  gtggggtgct tgggatcagg gtgtggggat gacggcgggg ctgacggagc
      86751  gggatgtccg tcgtgctgct gagtcgggta tgccgttgtt gacggttgat
      86801  caggtgtggg cgttgttcga tgcggcgttg gcgacggga gtgccgcgtt
  35  86851  ggtgccggtc cgtctggacc tggccgcact gcgcacccgg ggcgacatcg
      86901  caccgctcct ccgcggcctc gtccgcgcac cgctgcgccg caccgcggcc
      86951  accggcctcg ccaccggcgc ggacaccggc ctcgtccaac ggctcggccg
```

```
        87001  actcgaccac gcccaacgcc acgaggcact gctcgacatg gtccgcagca
        87051  gcgccgcgct cgtcctcggc cacgccgacg gcaacgccat cgacgccgaa
        87101  cgcgccttcc gcgacctcgg cttcgactcg ctcaccgcgg tcgaactccg
        87151  caaccgtctg cgcaccgcca ccggcctgca cctgtcggcc accatggtct
     5  87201  tcgaccaccc caccctgtcc gccctcgcgg agcacctgcg ggacgagttg
        87251  ttcggcgcgg tcgagagcga ggtgcgggtg ccggtccagg cactgccgcc
        87301  gaccgccgac gatcccatcg tggtggtggg catggcctgc cgtttccccg
        87351  gtggtgtgac ctcgcccgag gacctgtggc gcctggtcga cgacggcacc
        87401  gacgccatca ccaccttccc gaccaaccgc ggctgggacc tggacaacct
    10  87451  ctacgacccg gaccccgagc acttcggcac gtcgtacacc cgctccggtg
        87501  gcttcctgca cgaggcgggg gagttcgacc cggcgttctt cggaatgagc
        87551  ccgcgtgagg cgctggcaac cgactcccaa cagcgtctcc tgctggaatc
        87601  ctcctgggag gcgatcgagc gggccggcat cgaccgctg accctgcgcg
        87651  gcagcgccac cggcgtcttc gccggcgtga tgtacagcga ctacgggagc
    15  87701  atcctcggcg gcaaggagtt cgagggcttc caaggccagg gaagtgcggg
        87751  cagcgtggcc tcgggccgcg tctcctacgc cctcggcttc gagggcccgg
        87801  ccgtcacggt ggacacggct tgctcttcct ctctggtcgc cctgcactgg
        87851  gcggcgcagg cgttgcgggc gggggagtgc tcgttggcgt tggccggtgg
        87901  tgtgacggtg atgtcgacgc cgagcacgtt cgtggagttc tcgcggcagc
    20  87951  ggggtctggc gcctgatggt cgttccaagg cgttcgccga ggccgcggac
        88001  ggcgtcggct ggtccgaggg cgtcggcatc ctcgtcctgg agcgccagtc
        88051  ggacgcggtg cgcaacggcc acgagatcct cgccgtgatc cgcggctcgg
        88101  cggtcaacca ggacggtgcg tccaacggcc tgaccgcgcc caacggcccg
        88151  tcccagcagc gcgtcatccg tcaggcgttg gccagtggcg gcctgtccac
    25  88201  ggccgacgtg gacgccgtcg aggcgcacgg cacgggtacg acgctcggtg
        88251  acccgatcga ggcccaggcg ctcctggcca cctacggccg tgaccgcgac
        88301  cccgagaacc ccctgtggct gggctccctg aagtccaaca tcgggcacac
        88351  ccaggcagcg gccggtgtcg ccggtgtcat caagatggtc atggcgatgc
        88401  ggcacggcgt gctgccgcag accctgcatg tcgacgcgcc gtcctcgcac
    30  88451  gtcgattgga gcgtcggcgc cgtcgaactg ctcaccgagc agaccgcctg
        88501  gccggagacc ggccgggtcc gtcgcgccgg tgtctcctcc ttcggcatca
        88551  gcggcaccaa cgcccacgtc atcgtggaac agccggcgct cgtcgaaagc
        88601  ccggccgcgg agccgagcgg acgcgaaccc ggcgtcgttc cgctgccgct
        88651  gtccggaaag tcccccgagg ccctgcgcga ccaggccgca cgcctgctgg
    35  88701  ccgggttggc ggagcggccc gcgctgcgcc cgctcgacct cggctactcg
        88751  ctggcgacga cccgttcggc gttcgaccac cgggcggtgg tgctcgccac
        88801  cgaccgcgcc gatgcggtcc gcgcgctgac ggcgctcgcc gccgccgacg
```

```
88851   cggatctctc cgccgtcgtc ggcgacaccc gcacgggtcg tcacgcggtg
88901   ctgttctcgg gtcagggctc gcaacgcctg ggcatggggc gtgagttgta
88951   cgagcgtttc ccggtcttcg ccgaggctct cgatgtcgcg atcgaccacc
89001   tggacgccgc cttgcccgcc caggccagtc tgcgtgaggt gatgtggggc
89051   gacgatgtcg agctgctgga cgagacgggt tggacgcagc cggctctgtt
89101   cgccgtcgag gtggccctgt tccggctggt ggagagttgg ggtgtccgtc
89151   cggacttcgt ggccggtcat tccatcggtg agatcgcggc ggcgcatgtc
89201   gtcggggtgt tctcgctgga ggacgcctgc cgtctggtgg ccgcccgtgc
89251   gacgctgatg caggcgctgc cgaccggcgg cgcgatgatc gcgatccagg
89301   ccgccgagga cgaagtcacc cagcacctga ctgacgacgt ctcgatcgcc
89351   gccgtcaacg gcccgacctc cgtggtcgtc tccggggccg agagcgctgc
89401   ccgcacggtg gccgaccggc tcgcggagaa cggccgcaag acgacccggc
89451   tgcgggtctc gcatgcgttc cactcgccgt tgatggatcc gatgctggcg
89501   gagttccgtg cggtggccga gggcctgtcc tacgccaccc cgaccctccc
89551   cgtcgtctcg aacctgacgg gccggctggc cacggccgat gacctctgct
89601   cggccgagta ctgggcgcgc cacgtccgcg aggcggtccg cttcgccgac
89651   ggcgtcagca ccctggagaa cgagggcgtc accacgttcc tggaactggg
89701   accggacggc gtgctgtccg ccatggccca gcagtcgctc accggcgacg
89751   ccgccaccgt cccggccctc cgcaaggacc gcgacgagga gacgtccgcg
89801   ctcaccgccc tcgcccacct ccacacggca ggtctccgcg tcgactgggc
89851   ggcgttcttc gccggcagcg gcgccacccg cgtggacctg ccgacctacg
89901   ccttccagca cgccacctac tggcccaccg gcaccctgcc caccgcgcac
89951   gccgcggccg tcggcctcac cgccgccgag cacccgctgc tgaacggttc
90001   ggtcgaactc gccgaaggcg aagggggtgtt gttcaccgga cggctgtcac
90051   tgcagtcaca tccgtggctg gccgaccacg ccgtcatggg acaggtcctg
90101   ctgcccggca ccgcactgct ggaactggcg ttccgggccg gcgacgaggc
90151   cggttgcgac cgcgtcgagg aactgacgct cgccgcaccg ctcgtcctgc
90201   ccgagcgcgg tgcggtacag acccaggtcc gggtcggcgt cgccgacgac
90251   accggccgcc gtaccgtcac cgtccactcc cggcccgagc acgcgaccga
90301   cgtgtcgtgg acccagcacg cgaccggcac cctgaccatg ggctccgccc
90351   cggccgacac cggtttcgac gccactgcct ggccgcccgc cgacgccgaa
90401   cccctcgcca ccgacgactg ctacgcgcgc ttcacgacgc tcggcttcgc
90451   ctacgggccg gtcttccagg gcctgcgggc cgcctggcgc gccggtgacg
90501   tgctgtacgc cgaggtggcc ctggcggagt ccaccggcga cgaggcgacc
90551   gccttcggtc tgcaccccgc actgctcgac gccgccctgc acgcctccct
90601   cgtcgcccac gagggcgagg agagcaacgg cggactgccg ttctcctggg
90651   agggcgcgac cctctacgcg accggcgcca ccgcgctgcg cgtccggctg
```

```
        90701   accccgacgg gcaccgacgg ccgttcggtg gccatcgccg tggccgacac
        90751   cgccggtcgt ccggtcgccg ccatcgacaa cctcgtctcg cgccgggtct
        90801   ccggcgacca gttgaccggc gccgcgggac tggcccgcga cgccctgttc
        90851   accctggact ggaaccccgt accggagaac ctcgtaccgg agaaccctgt
5       90901   accggagaac accggcgggg gccacgccca ggaccaggac ggccggcccg
        90951   ccgcggccac cgtcgcgctg gtcggcgcgg acggcaccgc gatcgccgcc
        91001   gacctgaccg ccgccggcat ccacaccacc ctccaccccg acctcaccac
        91051   cctcgccacg accgacgccg acgttccgaa gacggtcctc atccccctca
        91101   ccggaaccgg aaccggaacc ggcaccggga ctgagtcgac ggacggaatc
10      91151   gggacggggg ccgccgagtc ggacgcgtcc gccccctccc cggccgaggt
        91201   cgcccacacc ctgtccaccg ccgcactcgc cctcgtccag gagtggaccg
        91251   cacaggagcg cttcgccggc tcccgcctgg cgttcgtcac gaccggggcg
        91301   acggccgccg gcggtaccga cgtcatggac gtggccgccg ccgcggtctg
        91351   gggcctggtc cgatccgccc agtccgaagc cccggacacc ttcgtcctga
15      91401   tcgaccgtga cccccggcccg gccggcacgc acgaccgcac agccgccgcc
        91451   gaacggggcc aactgctcct acgggcactg cacaccgacg aaccgcagct
        91501   cgccctgcgt gacggcggcg tgctcgccgc ccgcctggcc cgcttcgaca
        91551   ccgcggccgc gctcacccccg ccggccgacc gggcctggcg gctcgacagc
        91601   acggccaagg gcagcctcaa cggcctcgcc ctgacccccgt atccggcggc
20      91651   actggcgccg ctcaccggcc acgaggtgcg ggtcgaggtg cgtgccgcgg
        91701   gcctgaactt ccgtgacgtg ctcaacgcgt tggggatgta tccgggtgat
        91751   gatgtcggat cgttcggttc ggaggcggcc ggtgtggtcg tcgaggtcgg
        91801   accggaggtg accggcctgg cccccggcga ccaggtcatg ggcatgatca
        91851   ccggcagctt cggctcgctc gccgtggacg acgcgcggcg cctcgcccgc
25      91901   ctgcccgagg actggtcctg ggagacgggt gcgtcggtgc cgttggtgtt
        91951   cctcaccgcg tactacgccc tgaaggagtt gggtggtctg cgggcggggg
        92001   agaaggtgct ggtgcatgcc ggtgccggtg gtgtcggtat ggcggcgatc
        92051   cagatcgccc ggcatgtcgg tgccgaggtg ttcgccacgg ccagtgaggg
        92101   caagtgggac gtgttgcgct cgctcggcgt ggccgacgac cacatcgcct
30      92151   cctcccgcac cctcgacttc gaggcggcct tcgccgaagt cgccggcgac
        92201   cgcggcctgg acgtcgtcct caactccctc gccggtgact tcgtcgacgc
        92251   ctcgatgcgg ttgctcggcg acggcggccg gttcctggag atgggcaaga
        92301   ccgacatccg cgccgcggac tccgttcccg acggcctctc ctaccagtcc
        92351   ttcgacctcg cctgggtggt gccggaaacc atcggcacca tgctggccga
35      92401   gctgatggac ctcttccgca ccggcgcact gcggccactc ccggtccgca
        92451   cctgggacgt ccggcacgcc aaggacgcgt tccgcttcat gagcatggcc
        92501   aagcacatcg gcaagatcgt gctcacccctg ccccgctcct ggaagcccga
```

```
        92551  gggaacggtg ttggtgacgg gtggtacggg tggtctgggt gggttggtgg

        92601  cccggcatct ggtgcggtcg tgtggggtgc ggcatttgtt gttgaccagt

        92651  cgttctggtg tgggtgctgc gggtgcggcc gggttggtcg cggagttgga

        92701  gtcgttgggc gcgcgggttg tggttgcggc gtgtgatgtg ggtgatggct

    5   92751  cggctgttgc ggagttggtt gccggtgtgt cggagtcgta tccgttgtct

        92801  gcggtggtgc atgcggctgg tgtgttggat gacggtgtgg tgggttcgtt

        92851  gacgccggag cggttggctg cggtgttgcg tccgaaggtg gatggtgcgt

        92901  ggaacctgca tgaggcgacg cgtggtctgg atctggacgc gtttgttgtc

        92951  ttctcgtctg ttgcgggtgt gttcggggggt gcgggtcagg ccaactatgc

   10   93001  ggcgggtaat gcgtttttgg acgcgttgat ggttcatcgg gtggctggtg

        93051  ggttgcctgg tgtgtcgttg gcgtggggtg cttgggatca gggtgtgggg

        93101  atgacggcgg ggctgacgga gcgggatgtc cgtcgtgctg ctgagtcggg

        93151  tatgccgttg ttgacggttg atcagggtgt ggcgttgttc gatgcggcgt

        93201  tggcgacggg gagtgccgcg ttggtgccgg tccgtctgga cctggccgca

   15   93251  ctgcgcaccc ggggcgacat cgcaccgctc ctccgcggcc tggtcaaggc

        93301  gcccatccgc cgcgcagccg ccaccacacc cggcgacacc ggactcgccg

        93351  agcagctcac ccggctccag cgcgccgagc gacgggacac cctcctcgcg

        93401  ctcgtccgcg accaggccgc gatggtcctc ggccacacct cgggcgacgg

        93451  cgtcgacccg tcccgcgcct tccgcgacct cggcttcgac tcgctcaccg

   20   93501  cggtcgaact ccgcaaccgc atcggcgcgg ccaccggcct gcggctaccg

        93551  gccacggccg tcttcgacta ccccaccgcc gatgccctcg ccgcacacct

        93601  gctcaccgaa ctgctcggcc ccgacgccga gtcggacccc gacgagcccg

        93651  gcgaccccac cgcgggaccg accgacgacc ccatcgtcat catcggcatg

        93701  agctgccgct tccccggcga catcggctcg ccggaggacc tgtggcgcct

   25   93751  gctcggcgac ggcgccgacg tcgtcaccga cttcccgacc aaccgcggct

        93801  gggacctgga caacctctac gaccccgacc ccgcgcacgc cggcacctcg

        93851  tacgcccgca ccggcggttt cctgcacgac gccgccgact tcgacgccga

        93901  cttcttcggc atgagccccc gcgaggccat ggccacggac tcccagcagc

        93951  gcctgctgtt ggagtcctcg tgggaggcga tcgagcgggc cggcatcgac

   30   94001  ccgctgaccc tgcgcgacag ccgcaccggc gtcttcgccg gcgtcatgta

        94051  cagcggctac ggcacccgcc tcgacggcgc cgaattcgaa ggcttccagg

        94101  ggcagggcag cgcactgagc gtggcctccg gccgggtctc ctacacccttc

        94151  ggcttcgaag gcccggccat gacggtcgac accgcctgct cctcctcgct

        94201  ggtcgccctg cacctcgccg cacaggcact ccgcggcggt gagtgcaccc

   35   94251  tcgccctcgc cggtggtgtc accgtgatgt ccatcccgga caccttcatc

        94301  gagttctccc ggcagcgcgg actggccccc gacgccgct ccaagccgtt

        94351  ctccgagtcc gccgacggcg tcggctggtc cgagggcgtc ggaatgctgc
```

```
        94401  tcctggagcg ccagtcggac gccgtgcgca acggccacca gatcctggcc
        94451  gtggtgcgcg gctcggcggt caaccaggac ggtgcgtcca acggcctgac
        94501  cgcgcccaac ggcccgtccc agcagcgggt gatccgtcag gcgttggcca
        94551  gcggcggcct gtccacggcc gacgtggacg ccgtcgaggc gcacggcacg
  5     94601  ggcaccacgc tcggtgaccc gatcgaggcc caggccctcc tggccaccta
        94651  cggccgcgac cgcgaccccg agaacccgct gctgctcggt tcgatcaagt
        94701  ccaacctcgg ccacacccag gcagccgccg gtgtcgccgg cgtcatcaag
        94751  atggtcatgg cgatgcggca cggcgtgctg ccccgcagcc tgaacatcac
        94801  cgagccgtcc tcgcacgtcg attggagcgc cggcgccgtc gaactgctca
  10    94851  ccgagcagac cgcctggccg gagaccggcc gggcccgtcg cgccggtatc
        94901  tcctccttcg gcatcagcgg caccaacgcc cacgtcatcc tggagcagcc
        94951  ggaggccgcg cggcactcgg cgccggaaga agccgacacg gcggaggcag
        95001  ccgccaaggc gccggccacc gcgcacctgc ccgtaatgcc gtgggcactg
        95051  tccggcaaga cgccggaggc cctgcgtgcc caggccgcac gcctcctcgc
  15    95101  ccacctccag cagcgccccg aactcgcacc cgccgacatc gccctgtccc
        95151  tcgccaccca gcgctcccag ttcacccacc gggcagtcgt cctgagcacc
        95201  gaccgtgacg aggcgacccg cgcgctgtcc gccctcgcca ccaccgccgc
        95251  gtccgacccc tcggccctca ccggcacggt caccatggga cgttgcgcgg
        95301  tgctgttctc gggtcagggc tcgcaacgtc tgggcatggg gcgtgagttg
  20    95351  tacgagcgtt tcccggtctt cgccgaggct ctcgatgtcg tgatcgatca
        95401  cctggacgcc gccttgcccg cccaggccgg tttgcgtgag gtgatgtggg
        95451  gcgacgatgt cgagttgctg aacgagacgg gttggaccca gcccgcgctc
        95501  ttcgccatcg aggtggcgct gtttcggctg gtggagagtt ggggtgtccg
        95551  tccggacttc gtggccggtc attccatcgg tgagatcgcg gcggcgcatg
  25    95601  tcgtcggggt gttctcgttg gaggacgcgt gccgtctggt ggccgcgcgg
        95651  gcgacgctga tgcaggcgtt gccggccggt ggcgcgatga tcgcggtcca
        95701  ggcgaccgag gacgaagtca tcccgcacct gaccgacgag gtggcgatcg
        95751  cggccgtcaa cggcccgacc tccgtggtga tctcgggcgc agaagaggcc
        95801  acgcagaccg tggcacaaca cttcgccgac caggggcgcc ggacgaccgc
  30    95851  gctgcgggtc tcgcatgcgt tccactcgcc gctgatgatg ctggcggagt
        95901  tccgtgcggt ggccgagggc ctgtcctacg ccaccccgac cctccccgtc
        95951  gtctcgaatc tgacgggcca ggtggccacg gccgacgaac tctgctcggc
        96001  cgagtactgg gtgcgccacg tccgtgaggc ggtccgcttc gccgacggtg
        96051  tgacggccct cgaagccgag ggcgtgcgga ccttcctgga actcggcccg
  35    96101  gacggcgtcc tcgccgccat ggccagggaa accgtcgccg acgacacggt
        96151  caccgtcccc gtcctccgca ggaacatgcc cgaggaacgg accctgctca
        96201  ccgcactcgg ccggctccac accaccggaa ccccgatcga ctgggccgcc
```

```
        96251  ctcctggccc cgaccggcgc ccgcccggtg gacctgccga catacgcgtt
        96301  ccaacaccgt cccttctggc cctccggccc ccgcgacacc gcggatgccg
        96351  ccgccgtcgg catcgccggc gcgagccacc cgctcctcaa cggcatcgtc
        96401  gaactcgccg acgaagaggg cctgttgttc accggacggc tgtcactgca
     5  96451  gtcgcatccg tggctggccg accacgccgt catgggacag gtcctgctgc
        96501  ccggcaccgc actgctggaa ctcgccctgc gcgccggcga cgaggtcggc
        96551  tgtgaccacg tcgaggaact gacgctcgcc gcaccgctcg tcctgcccga
        96601  gcgcggcgcg gtacagaccc aggtccgggt cggcgtcgcc gacaccaccg
        96651  ggcgccgcac cgtcacgatc cactcgcgtc ccgcacgcgc cacgaccacc
    10  96701  gacagtgaca cccacaccgg caccgacacc ccgtggaccc aacacgccac
        96751  cggcgtcctc gtcgccggcc tgccggcgac ggcaaccgtc ccgttcgatg
        96801  ccaccgtgtg gccgccggcg cacgccgaac ccgttgacct ggcggacttc
        96851  tacgcgtccc gggccggcga aggattcggc tacgggcccg ctttccaggg
        96901  cctgcgagcg gcctggcgcc gcgacggcga ggtgttcgcc gatgtcgcac
    15  96951  tgccggaggc cggccgtacc gaagccgagg cgtacgggct gcatccggca
        97001  ctgctcgacg ccggactgca cgcagcctgg ctcgtcgccc cggacggggga
        97051  gcccacacgg acgggcagcg tgccgttctc gtggcgcggc gtcttcctgg
        97101  ccgcttccgg tgcctcctcg gtccgcgtcc gactcggccg cgactccgac
        97151  ggaacgctga gcctggccat cgccgacacc accggtgcac cggtcgcgtc
    20  97201  cgtacaggcc ctctccatgc gcaccgtctc ggtgacggcc ctgagcgcca
        97251  ccgcgggcct cgcccgcgac gcgctgttcc gcctggactg ggcctcggcc
        97301  ccggagccgg cgtgccagcc ggacgacacg gtgaccgtga tcccggcggt
        97351  cgcggtcgtc ggtacggaaa cctccgaact cacctccgag ctcaccgcgg
        97401  ccctgcgtgc ggcgggcgcc gacgtcgacg tccgcacgac cctgtcgacc
    25  97451  gatgaacccg cgcccgcgct gatcgcgctg ccgctcgtcg cctccgacca
        97501  gaccggcacc gcagaagcgg cacccgtccc ggcggccgtg cacgacctca
        97551  cccgccgagc cctggccctc gtacagaccc gcctgcaaga gcagcacttc
        97601  gcggacacga agttcgtctt cgtgacccgt ggtgcgacgg tcgggcgtga
        97651  tgtggctgct gctgcggtgt ggggtctggt gcgttcggcg cagtcggaga
    30  97701  atccggggttg ttttgctctg gtcgatctgg atccggatgg tgcggtgggt
        97751  gcggctgcgc tcgtcgctgc gttggtcagt ggtgagccgc agcttgcggt
        97801  gcgcggtgat gtgttgcggg tcgcgcgtct ggtgcggcgg ccgctcaccg
        97851  aggtcggtgc gggtgctgat ggcaccgggg atggcgtcgg gggtggctct
        97901  ggtgtgtcgt tctcgggtga gggtgcggtc ctggtcactg gtggtacggg
    35  97951  tggtctgggt gcggtgttgg cgcgtcatct ggtggccgag tatggggtgc
        98001  gggatctgct gttggtcagt cgcagtggtg aacgtgccgt gggtgctggg
        98051  gagttggtgg cggagcttgc gggtgtgggt gcgcgggtgc gggtggttgc
```

```
        98101  gtgtgatgtg  accgatcgtg  ccgcggtggt  ggagttggtt  ggcgggcatg
        98151  cggtgtccgc  ggtggttcat  gcggctggtg  tgctggatga  cggcatggtg
        98201  ggtgcgttga  ccggggagcg  gttgtccgcg  gtgctgcggc  cgaaggtgga
        98251  tgctgtctgg  catctgcatg  aggcgacccg  cggcctggat  ctggacgcgt
   5    98301  tcgtcgtctt  ctcctctctc  gccggggtct  tcggcagtcc  cggccaggcc
        98351  aactacgcgg  ccgcgaacgc  cttcctggac  gcgctgatga  cgcggcgccg
        98401  ggcggaggga  ctgcccggcc  tgtcactcgc  atggggaccg  tggtcgctga
        98451  ccgatggcac  atcgggcatg  ctcgcggacg  ccgaggccga  tcgcctgacc
        98501  cgttcgggag  tgccaccgct  gaccgcggag  caaggactgg  cactgttcga
   10   98551  cgcggccctg  gcgaccggtg  acgccacctg  cgtcccggtc  cgcctggacc
        98601  tgtcggcact  gcgtgcccag  ggtgaggtgc  cgcccttgct  gcggtccctg
        98651  atccgaggcc  gctcgcgccg  cgccgccgcc  gcggaatcgg  caaccgccac
        98701  cgggctgcgg  gaacggctcg  tcggactgaa  cccggtcgag  cgacaggaag
        98751  tcctcctgga  cctcgtgcgc  ggccaggtgg  ccctggtcct  cggccacgcc
   15   98801  gacgccgacg  acgttcatcc  ggcacgtgcc  ttcagggagt  tgggcttcga
        98851  ctccctcacc  tcggtggaac  tgcgcaaccg  cctcaacacc  gtcaccggtc
        98901  tgcggctccc  cgcgaccatg  gtgttcgact  atccgaccgt  cgaagtgctg
        98951  gtctcctacg  tcctggacga  gttgctgggc  acggatgccg  aggtggcgac
        99001  cgtgcagccg  gccgcggttg  cggtggcgga  cgatccgatc  gtcatcgtgg
   20   99051  gcatggcctg  ccgctacccc  ggcgggggtcg  cctctccgga  cgacctgtgg
        99101  aggctcgtca  cggacggcgt  cgacgcggtg  tcccccttcc  cgaccaaccg
        99151  tggttgggac  gtcgaatccc  tctatcaccc  ggaccccgac  catctcggta
        99201  cctcctacac  gcgctcgggt  gggttcctgc  atgaggcggg  ggagttcgat
        99251  ccgggggttct  tcgggatgag  tccgcgggag  gcgttggcga  ccgattccca
   25   99301  gcagcggttg  ttgttggagt  cgtcgtggga  ggcgatcgag  cgggccggta
        99351  ttgatccggt  gagtttgcgg  ggtagtcgga  cgggtgtgtt  cgcgggggtg
        99401  atgtacagcg  attacagcgc  gatgttggcg  agtccggagt  tcgagggttt
        99451  ccagggcagt  gggagttcgc  cgagtttggc  gtcgggtcgg  gtggcctaca
   30   99501  cgttggggtt  ggaaggcccg  gcggtgacgg  tggatacggc  gtgttcgtcg
        99551  tcgttggtgg  cgatgcactg  ggcgatgcag  gcgttgcgta  gtggtgagtg
        99601  tgggttggcg  ttggccggtg  gtgtgacggt  gatgtcgacg  cctgcggtgt
        99651  ttgtggactt  tgctcggcag  cggggtttgt  cgccggatgg  tcggtgcaag
        99701  gcgtttgcgg  atgcggccga  tggtgtgggc  tggtccgagg  gcgtcggcgt
        99751  gttggtcctg  gagcggcagt  cggacgcggt  gcgcaatggt  cacgagattt
   35   99801  tggctgtggt  gcggggttcg  gcggtcaacc  aggatggtgc  gtccaatggt
        99851  ttgacgcgc  ccaatggtcc  gtcgcagcag  cgggtgatcc  ggcaggcgtt
        99901  ggccagtggt  ggtctgacgg  ccggtgacgt  ggatgtggtg  gaggcgcatg
```

SUBSTITUTE SHEET (RULE 26)

```
        99951  gtacgggtac gacgctcggt gatccgatcg aggcgcaggc gttgttggcg
       100001  acgtatgggc gggatcgtga gcctgagcgg ccgttgttgt tgggttcggt
       100051  gaagtcgaat ctggggcata cgcaggctgc tgcgggtgtg gcgggcgtta
       100101  tcaagatggt gttggcgatg cggcatggtg tggtgccgcg gacgttgcat
    5  100151  gtggatgcgc cttcttcgca tgtggactgg tccgagggtg cggtggagct
       100201  gctcagtgag caggcggctt ggccggagac gggtcgggtg cggcgggcgg
       100251  gtgtctcctc cttcggcatc agcggcacca acgcgcacgt gattctggag
       100301  cgtccggagg ccgcgcggcg tccggtgatg gagacgaaca ccgtggagcc
       100351  gtccacggtg ccgtgggtgc tttccggcaa gacaccggag gccctgcgcg
   10  100401  cccaggcggc gaaactgttg tcgtcgatcg aggaacgccc ggagcttcgc
       100451  ctcgtcgacg tgggaatgtc cctggtcacc ggccgctcga cctttgaaca
       100501  ccgtgcggtg gtactcgccg ccgaccgtgc cgacgccgcc cgcgcattgt
       100551  cggcgattgc tgccgacgag gccgatgctg ccgctgccac cggccgcgtc
       100601  ggcgcgggtc gtcacgcggt gctgttctcg ggtcaggtg ctcaacgtct
   15  100651  gggcatgggg cgtgagttgt acgagcgttt cccggtcttc gccgaggctc
       100701  tcgatgtcgt ggtcgaccac ctggacgccg ccttgcccgc ccaggccggt
       100751  ttgcgtgagg tgatgtgggg cgacgatgcg gagttgctga acgagacggg
       100801  ttggacgcag ccggctctgt tcgccatcga ggtggcgctg ttccggctgg
       100851  tggagagttg gggtgtccgt ccggacttcg tggccggtca ttccatcggt
   20  100901  gagatcgcgg cggcgcatgt cgccggggtg ttctcgctgg aggacgcctg
       100951  ccgcctggtg gccgcccgtg cgacgctgat gcaggcgttg ccggccggcg
       101001  gtgcgatgat cgcggtccag gcgaccgagg acgaagtcac cccgcatctg
       101051  accgacgacg tagcgatcgc cgccatcaac gggccgaacg cactggtcgt
       101101  atcgggtgtg gaggatgccg ccgtcgagat cggggcgcgg ttcgcggccg
   25  101151  aggggcgtcg cacgacccga ctccatgtgt cgcatgcgtt ccactcgccg
       101201  ttgatggatc cgatgctggc ggagttccgt gtggtggcgg agggcctgtc
       101251  ctatgctgct ccgtccctcc ccgtcgtctc gaatctgacg ggccaggtgg
       101301  ccacggccga cgaactgtgc tcggccgagt actgggtgcg ccacgtccgc
       101351  gaggcggtcc gcttcgccga cggggtgacg gccctcgaag ccgagggcgt
   30  101401  gcggaccttc ctggaactcg gcccggacgg cgtcctcgcc gccatggcag
       101451  gagcctcgct caccgaatcc tccctcgcgg taccgctgct ccgtaaggac
       101501  cggccggagg aaccggcggc actcgccgcc ctggcccagt tgcacatcgc
       101551  cggcgcgcgc gtcgattggc ccgtgctctt cgctggtgtg ggtgcggggc
       101601  gggtggagtt gccgacgtat gcgttccagc gtgggtggtt ctggccggtt
   35  101651  ggtcgggttg gtgttggtgg tgatgtgggt gctgtggggc ttgggtctgc
       101701  ggggcatccg ttgttgggtg ctgcggtgga gttggctgcg ggtgcggggg
       101751  tggtgttgac gggtcgtctg tcgttgtcgt cgcatggttg gttggctgat
```

```
        101801   catgcggtga tggggcgggt gtttgttcct ggtacggcgt tgctggagat
        101851   ggtgatgcgt gctggtgatg aggtggggtg tggtcgtgtt gaggagctga
        101901   cgttggcggc gccgttggtg ttgcctgagc gtggtggggt gcgggttcag
        101951   gttgctgtgg atgctcctga tgctgcgggt cgtcgtggtg tggggggtgta
5       102001   ttcgtgtcct gatggtgtgg gtcaggcggt gtggtcgcag catgctgtcg
        102051   gtgtgttggc ctctggtgtg gctgaccagg tcggtgggtt cggtgacggt
        102101   ggtgtgtggc cgccgcaggg tgcggtgtcg gtggatgctg agggctgcta
        102151   cgagctgttt gcggatgctg ggttcggtta tggcccggtg ttccaggggt
        102201   tgcgtgcggt gtggcgtcgc ggcgaggaac tcttcgcaga ggtcgccctg
10      102251   tcggacgagg ttgctgagag cgctgatacg gcgaccggtt tcgggttgca
        102301   cccggcgttg ctggatgcct cgctccatgc ctcgcttctc tcctcccttg
        102351   aaggtcaatc cgccgatggt gggcctgcgt tgccgttcgc gtgggagggt
        102401   gtttccctct tcgcctcggg tgcgacggct ttgcgcgtgc ggttggcgcc
        102451   ggcgggtgag catgcggtgt cggtgaccgc ggtggatccg accggtgcgc
15      102501   cggtgatttc catcgacgcg cttcgtaccc gtcgcctcac cctcgatgag
        102551   gtcaacgcat ctcacaccca gctgagcgat gcgctcttcg gcgtccaatg
        102601   gaccacggtc ccgagcaccc cggccgccga ccacccgtcg gtagccatca
        102651   tcggaaccga ccacctgggg ctcgccgaag cgctcagcag ttcctctgct
        102701   ggcgcgacca cgaccacgac cgcggccgcg tacgagagcc ttgacgcgct
20      102751   gatcgcggcg gggcccgaag tgtccgtccc tgacgtcaca ctcatcggtc
        102801   tcaccaccga ggacgccatc gctcagtacg tgaacgacca cgacgccacg
        102851   gtggccggcc aaggcaccat cggagccggc gcggcggccg tagacgcggc
        102901   tcgccgcctc accgccgagg ccctgcgcac gatccaggca tggttggccg
        102951   acgagcgtct ggcggcaagg cgcctggtct tcgtgacccg cggcgcggcc
25      103001   gacgggcagg atgtcgccgc ggcggccgtc cagggcctgg tgcgctccgc
        103051   ccagaccgag aacccgggca ccttcggcct cctcgacctc gacgggaccg
        103101   aggcgtcgac cgcggtcctc ggcgaggctc tcacctccga cgaaccgcaa
        103151   ctgcttctgc gcgacggaca cctgcacgcc gccaggctga cccgcctggc
        103201   gtcgcccgcc gacacggccg tgcccacgga gtggaacgcg gacggcacgg
30      103251   tgctgatcac cggtggtacc ggcgggctcg gcgcgcagtt cgcacggcac
        103301   ctcgtcgaca ggtacggcgt ccgcaatctc ctgctcgtca gccggcgcgg
        103351   ccccgatgcc ccgggaacca cggagttggt cgccgagctg acggcgcacg
        103401   gtgccgaggt ggccgtgcag gcatgtgacg tggccgatgg cgatgcggtg
        103451   gcggcgttgg tcgccggcgt gccggatgag cacccgctga gggcggtcgt
35      103501   ccacacggcc ggtgtgttgg acgacggagt gatcggctcg ctcaccgagg
        103551   agcggctcgc caccgtcctg cggcccaagg cggatgccgc ctggcatctg
        103601   cacgaagcga cccgcggcct ggacctggac gcgttcgtcg tcttctcctc
```

	103651	cgtcgcaggg	gtcttcggcg	gcgccggcca	ggccaactac	gccgcggcca
	103701	acgccttcct	ggatgcgctg	atggcccagc	gccgggcagc	gggcctgccc
	103751	ggactctcgt	tggcctgggg	gccgtgggac	cagaccggcg	gaatgacggg
	103801	catgctgtcg	gacgccgagg	ccgaccgcct	cgcccgctcc	ggcatcccgc
5	103851	cgctctccgc	ggagcagggc	ctcgccctct	tcgacgcggc	actcgctctt
	103901	gccggaacca	gcacgccaga	cagggcagcc	ggcagcgccg	ccgccagcac
	103951	gtcgggaacc	ggcgacacga	tcgccatccc	ggccgcggcc	ctcgtcgcac
	104001	cggtccggct	cgacctggcg	gcgctggccg	cgcagggtga	ggtccccgcg
	104051	attctgcgcg	gactggtgcg	gacccgcacc	cgccgtacgg	cggccggcgg
10	104101	ttccgtcacc	gtggccggac	tcgtcaaccg	cctgtccggg	ctcaccgccg
	104151	acgagcggcg	ccaggaactt	ctcgaactgg	tccgcactca	ggcagccctc
	104201	gtcctcgggc	acgccgatcc	ggcgtccgtg	gactccaccg	cacagttccg
	104251	tgacctcggc	ttcgactcgc	tgaccgccgt	cgagctgcgc	aaccggctga
	104301	gcacggccac	cggcctgcgc	ctgaccgcaa	ccctggtctt	cgactacccg
15	104351	aacaccgatg	ccctcgcgga	gcacctgcgg	gacgagttgt	tcggcgcggt
	104401	cgagagcgag	gtgcgggtgc	cggtccaggc	actgccgccg	accgccgacg
	104451	atcccatcgt	ggtggtgggc	atggcctgcc	gtttccccgg	tggtgtgacc
	104501	tcgcccgagg	acctgtggcg	cctggtcgac	gccggcaccg	acgccatcac
	104551	caccttcccg	accaaccgcg	gctgggacct	cgaatcgctc	tacgacccgg
20	104601	accccgcaca	cctcggcacc	tcctacaccc	gctccggtgg	cttcctgcac
	104651	gaggcggggg	agttcgaccc	ggcgttcttc	ggaatgagcc	cgcgtgaggc
	104701	gctggcaacc	gactcccaac	agcgtctcct	gctggaatcc	tcctgggagg
	104751	cgatcgagcg	ggccggcatc	gacccgctga	ccctgcgcgg	cagcgccacc
	104801	ggcgtcttcg	ccggcgtgat	gtacagcgac	tacgggagca	tcctcggcgg
25	104851	caaggagttc	gagggcttcc	aaggccaggg	aagtgcgggc	agcgtggcct
	104901	cgggccgcgt	ctcctacacc	ctcggcttcg	aaggtcccgc	cgtcaccgtg
	104951	gataccgcct	gctcctcctc	cctggtcgcc	ctgcacctgg	cagcccaggc
	105001	ccttcgggcg	ggtgagtgca	cgctcgcgct	cgccggtggt	gtgacggtga
	105051	tgtccacgcc	aggcacgttc	gtggagttct	cgcggcagcg	gggtctggcg
30	105101	cctgatggtc	gttccaaggc	gttcgccgag	gccgcggacg	gcgtcggctg
	105151	gtccgagggc	gtcggcatcc	tcgtcctgga	gcgccagtcg	gacgccgtgc
	105201	gcaacggcca	cgagatcctc	gccgtgatcc	gcggctcggc	ggtcaaccag
	105251	gacggtgcgt	ccaacggcct	gaccgcgccc	aacggcccgt	cccagcagcg
	105301	cgtcatccgt	caggcgttgg	ccagtggcgg	cctgtccacg	gccgacgtgg
35	105351	acgccgtcga	ggcgcacggc	acgggtacga	cgctcggtga	cccgatcgag
	105401	gcccaggcgc	tcctggccac	ctacggccgc	gaccgcgacc	ccgagaaccc
	105451	gctgctgctc	ggttcgatca	agtccaacct	cggccacacc	caggcagcgg

```
105501  ccggtgtcgc  cggtgtcatc  aagatggtca  tggcgatgcg  gcacggcgtg
105551  ctgccgcaga  ccctgcatgt  cgacgcgccg  tcctcgcacg  tcgattggag
105601  cgtcggcgcc  gtcgaactgc  tcaccgagca  gaccgtctgg  ccggagaccg
105651  gccgggtccg  tcgcgccggt  gtctcctcct  tcggcatcag  cggcaccaac
105701  gcccacgtca  tcctggaaca  gccggaggcc  gtgcagcgcc  tggcaccggg
105751  agcagcagag  accgtggagc  cggtcgccat  caagccgtcg  gcggaaccgt
105801  ccctggtgcc  gtgggcgctg  tccggcaagt  cacccgaggc  cctgcgcgcc
105851  caggccgcac  gcctccgtga  cttcctggcg  gaacggcccg  aaccgcgctc
105901  gatcgacatc  ggccactcac  tggccgtcac  acgctcgcag  ttcgaccacc
105951  gcgcgatcgt  gctggtcgac  gatgcgaagg  ccccggcgga  cagcctggcc
106001  gccctcgcgg  ccctggcctc  cggtgtggcc  gatcccgccg  tcgtctccga
106051  cgcggtatcg  accggcggtt  cggcagtgct  gttcacaggt  cagggtgctc
106101  aacgtctggg  catggggcgt  gagttgtacg  gccgtttccc  ggtcttcgcc
106151  gaggctctcg  atgtcgtggt  cgaccacctg  gacgccgcct  tgcccgccca
106201  ggccggtttg  cgtgaggtga  tgtggggcga  cgatgtcgag  ttgctgaacg
106251  agacgggttg  gacccagccc  gcgctcttcg  ccgtcgaggt  ggcgctgttc
106301  cggctggtgg  agcgttgggg  tgtccgtccg  gacttcgtgg  ccggtcattc
106351  catcggtgag  atcgcggcgg  cgcatgtcgc  cggggtgttc  tcgctggagg
106401  acgcctgccg  tctggtggcc  gcccgtgcga  cgctgatgca  ggcgctgccg
106451  accggcggcg  cgatgatcgc  ggtccaggcc  accgaggacg  aagtgacccc
106501  gcacctgacc  gacgaggtgg  cgatcgcggc  cgtcaacggc  ccgacctccg
106551  tggtgatctc  gggcgcagaa  gaggccacgc  agaccgtggc  acaacacttc
106601  gccgaccagg  ggcgccggac  gaccgcgctg  cgggtctcgc  atgcgttcca
106651  ctcgccgctg  atggatccga  tgctggcgga  gttccgtgcg  gtggcggaag
106701  gactgtccta  cgccaccccg  tccctccccg  tcgtctcgaa  tctgacgggc
106751  tggctggcca  cggccgacga  actgtgctcg  gccgagtact  gggtgcgcca
106801  cgtccgcgag  gcggtccgct  tcgccgacgg  catcaccacc  ctcgaagccg
106851  agggcgtgcg  gaccttcctg  gaactcggcc  cggacggcat  cctgtccgcg
106901  ctggctcagc  agtccctcgc  cggcgaagcc  gtcaccgtgc  ccgtcctgcg
106951  caaggaccgc  ggtgaggagt  ccacggccct  gacggcccga  gcgcatctcc
107001  acacccgcgg  actgatcgaa  gactggcagg  acttcttcgc  tggtgtgggt
107051  gcggggcggg  tggagttgcc  gacgtatgcg  ttccagcgtg  ggtggttctg
107101  gccggttggt  cgggttggtg  ttggtggtga  tgtgggtgct  gtggggcttg
107151  ggtctgcggg  gcatccgttg  ttgggtgctg  cggtggagtt  ggctgcgggt
107201  gcggggggtgg  tgttgacggg  tcgtctgtcg  ttgtcgtcgc  atggttggtt
107251  ggctgatcat  gcggtgatgg  ggcgggtgtt  tgctcctggt  acggcgttgc
107301  tggagatggt  gatgcgtgct  ggtgatgagg  tggggtgtgg  tcgtgttgag
```

```
107351   gaactgacgt tggcggcgcc gttggtgttg cctgagcgtg gtggggtgcg
107401   ggttcaggtt gctgtggatg ctcctgatgc tgcgggtcgt cgtggtgtgg
107451   gggtgtattc gtgtcctgat ggtgtgggtc aggcggtgtg gtcgcagcat
107501   gctgtcggtg tgttggcctc tggtgcggct gaccaggtcg gtgggttcgg
107551   tgacggtggt gtgtggccgc cgcagggtgc ggtgtcggtg gatgctgagg
107601   gctgctacga gctgtttgcg gatgctgggt tcggttatgg cccggtgttc
107651   caggggttgc gtgcggtgtg gcgtcgcggc gaggaactct tcgcagaggt
107701   cgccctgtcg gacgaggttg ctgagagcgc tgatacggcg accggtttcg
107751   ggttgcaccc ggcgttgctg gatgcctcgc tccatgcctc gcttctctcc
107801   tcccttgaag gtcaatccgc cgatggtggg cctgcgttgc cgttcgcgtg
107851   ggagggtgtt tccctcttcg cctcgggtgc gacggctttg cgcgtgcggt
107901   tggcgccggc gggtgagcat gcggtgtcgg tgaccgcggt ggatccgacc
107951   ggtgcgccgg tgatttccat cgacgcgctt cgtaccgtc gcctcaccct
108001   cgatgaggtc aacgcatccc acacccagct gagcgatgcg ctcttcggcg
108051   tccaatggac cacggtcccg agcaccccgg ccgccgacca cccgtcggta
108101   gccatcatcg gaaccgaccc cttcggcctc gcagacggcc tttcggacgc
108151   cttgcccctg gtcgaggagc gcggtgacct cgcggcgctc gcagcgtcgg
108201   agcacccggt accggacctc gtcctcgtcc cggtagcggg cacccggcgc
108251   acaggcgtac ctgcggacgc cgaaggacac accgacgccg ggacatccga
108301   catgctccga tccgtgcgtg aggccaccgc acaggtactg gagcagatcc
108351   agcagtggtt ggcggacgac cggttcgagg cggcgcggct ggtgttcgtg
108401   acgcgcgggg cggtttccgt gggtgagggc ggcatcgccg acctggcggc
108451   ctcggccgtc tggggtctgg tgcggtcggc gcagtcggag aatccgggct
108501   gcttcggtct tctcgacctc gacctcgacc tcgcccttga ctccgacctt
108551   gcccccgagg tcgacatcga gcgcgaccgt gaccgcgatc cggtcggtgg
108601   gaccgtgcag cccgcgctcg ccgcggccct gcacgcgacc gccgacgagc
108651   cgcagttggc actgcgcggc gggaccgtgc aggccgcccg actgacccga
108701   atccccgcgc cgcagaccga ccgtgccgag accgaccctg ccgagaccga
108751   ccgtccggag atcgacaccc ggcggcccgg cacggtgctc atcaccggtg
108801   gtaccggtgg cctcggtggg ttgctcgccc ggcacctcgt cgccgagcgg
108851   ggggtacgga gcctggtgct cgccagccgg agcggtctcg cggccgaggg
108901   agcggagaag ctggtcgccg acctcgaagc gctcggtgcc gtggtggccg
108951   tgcagacgtg tgatgtggcc gatggcgatg cggtggcggc gttggtcgcc
109001   ggcgtgtcgg acgagtaccc gctgacggcg gtcgtccaca cggccggtgt
109051   gttggacgac ggagtgatcg gctcgctcac cgaggagcgg ctcgccaccg
109101   tcctgcggcc caaggcggat ccgcctggc atctgcacga ggcgacccgc
109151   gatctggacc tggacgcgtt cgtcgtcttc tcctccctcg ccggcgtcct
```

```
109201  cggtggcgcc ggtcaggcca actacgcggc ggcgaacacg ttcctggacg
109251  ccttgatggc gcagcgtcgc gccgccgggt tgccgggtgt gtcgctggcg
109301  tggggtccgt gggaccgggc cggcggcatg acggggaccc tgtcggacgc
109351  cgaggccgac cgcctcgccc gctccggtgt tccgccgatc tcggcggagc
109401  agggccttgc gctgtacgac gcggcgaccg ccggtgagcg gccgctggtg
109451  gtgccggtgc ggctggacct cgccgcgctc cgcgggctcg gtgatgtccc
109501  ggcgctgctg cgcggactgg tccggacgcc cgcgcggcgg accgcggcgg
109551  ccggtgcggc gccgtcggcc gatgtgctca cccggcagtt ggccgggctc
109601  ggcggggcgg agcaggagga ggtcctgctg aggctggtgc gcggtcaggc
109651  cgcggtggtg ctcgggcacg ccgacggctc ggcgatcggt gcggggcgac
109701  agttccagga gttgggcttc gactcgctga ccgcggtgga gttccgcaac
109751  cgactcaacg cggccaccgg actgcggctg ccggccaccc tgctgttcga
109801  ctacccgacg ccggccgacg tcgtcgggca cctgcgcggc cggctcggca
109851  ccggggaggt gtcgggtgcg ggctcggtgc tggcggcgct ggacaacctt
109901  gaggcggtga tcgccgggct gtccctcgac gacgcggggg agcaccagtt
109951  ggtggccggc cggctggagg tcctcagggc gaagtgggcg gacatgcgaa
110001  gcgcggaggg agctgtggac ggcggtgcgg acgtcgacat cgaggaggcg
110051  tcggacgacg acatgttcgc gctgctggac gacgagctgg ggctgaactg
110101  agccgctccg catgagcagt tccgcaccag aagttccaca gtggcctggt
110151  ccacagtgac ctggtccaca gtgagctgtt ccacagcgac ctgttccgcg
110201  gcggttccgt agtgagccgt tccgcttagg cgggaaaccg gggccccggt
110251  ggtcggacgc ttgattccgg ccaccgggag cgtcaaaccg gcctcttcta
110301  aagaaaggga aagaactagg gaaacacccca cagcccgatc atgtgcaatg
110351  aattccgtgg ctcggaaaac cattcgcggg ccatggagtt acggtgtgat
110401  cacgtgtccg tacggcagtg acggcgtgca cgggttgcca cggcaggtcc
110451  ccggccgagg cgtctccccg ctgctgcccg gcggtgcgtg ccgcttggtt
110501  cgtccctagg aggtcccccc atgaccacgt ccaccgagga gagcctgtgg
110551  gcccggtgct tccatccagc accggccgcc cccgtccggc tcttctgttt
110601  cccacatgcg ggcggctcgg cctccttcta cttcccggtg tcggcccaac
110651  tgtcctcggt tgccgaggtg ttcgccatcc agtacccggg gcgccaggac
110701  cggcgcaagg aagccggtgt cagtgacctc gcgaccttgg ccgaccaggt
110751  ctacgacgcg ctgcgcccccc tgctgaagga gcggccgagc acgttcttcg
110801  ggcacagcat gggcgcgacg ctggccttcg aggtggcccg gcgcttcgag
110851  gccgacgacg gtgacctggt ccggctgttc gcctccgggc gccgggcccc
110901  ctcccgcgtg cgtgaagagg ccgtgcaccg gcggtccgac gacggcatcg
110951  tcgaggagct gaagctgctc gccggcacca acaccgcgct gctcggcgac
111001  gaggagatcc tgcggatgat cctgcccgcg atccgcagcg actaccaggc
```

```
       111051  catcgagacc  taccgctgcc  cgcccgacgt  caccgtccgg  gcgccgctga
       111101  ccgtcctcac  cggcgaccgc  gacccgaaga  cctccctgga  cgaggccgag
       111151  gcgtggcgcg  gccacaccac  cggggacttc  gacctcaagg  tgcttcccgg
       111201  tgggcacttc  ttcgtcagct  ccgaggcccc  ggcgatcatc  gatctgctcc
    5  111251  gggcgcacct  cgccggcaac  ggctagcggg  cgcactgcgg  caggccggcg
       111301  gtgccggtct  gccgcacggc  ccccgcgccg  cctgagacgg  caccatgcca
       111351  cccggcgacg  cgcgttgcgt  gcgcctgtgg  agcaccgtcc  cgcgtgcgta
       111401  cgcgggcgcg  cggaggacgt  gctcggccgg  ccggatgtcc  tggaggagac
       111451  cattcggaaa  acgtcaaatg  tctgacagct  gctgctcatt  tactccagcc
   10  111501  cgcactatat  gatcttgtgg  tggggcatgt  acggagcgtg  ccactcgttc
       111551  gctcctggtg  catccatccc  gaccgtccgg  atttgtgtgc  accggacggc
       111601  cgtggtcctg  gtacgcgtcg  gcgcgaggtc  ggctcctgtt  ctcaccccac
       111651  gtggaggaac  gcgtgatgcg  gaagcagtcg  ggttcttccg  gcttactgac
       111701  cacgctggtg  ggacgcgacg  acgaactgcg  caccctcgcc  cggcacgccg
   15  111751  cggccgcccg  cgacgggcgg  gccggcctgg  tcctgctcca  cgggcccgcc
       111801  ggcatgggca  agaccagtct  gctgcggtcc  ttcacggcga  gcgatgtctg
       111851  ccgcggcatg  acggtgctgt  acggcacctg  cggcgagacc  gtcgccggcg
       111901  ccgggtacgg  cggtgtgcgc  gaactcctcg  gcgggctcgg  cctgagcggc
       111951  ggcgacgccc  ggcgctcgcc  cctcctggag  ggcctggcgg  cccgcgcgct
   20  112001  gcccgcactc  accgcagacc  ccgccggtcc  cgacgccgcc  acgggtgcct
       112051  acccggtgct  gcacggcctg  tactggctcg  ccgcccgcct  catggcccaa
       112101  cggccgctgg  tcctcgtcct  cgacgacgtc  cactggtgcg  acgaacgctc
       112151  cctggcctgg  atcgacttcc  tgctgcgacg  cgccgaggac  ctgccgctgc
       112201  tggtcgtcct  ggcctggcgc  agcgaggccg  aaccggtcgc  gcccgcggtg
   25  112251  ctcgccgaca  tcgccgccca  gcgccgcccc  accgtgctcg  gcctgcaccc
       112301  gctcggcccc  gacgacatag  gcgaaatggt  gcgtcgcgtc  ttccggacca
       112351  cggccgcacc  gtcgttcgtc  agccgggtcg  ccgccgtgtc  cggcggcaat
       112401  ccgctggccc  tcgcccgcct  cctcgacgaa  ctccgcgccg  agggcgtccg
       112451  gccggacgcc  gccggggagc  gccgggccgc  cgaggtcggc  agtcacgtcc
   30  112501  tcgcccgctc  ggtgcgctgc  ctgttggagc  gccggccgcc  ctgggtgcgc
       112551  ggcgtggccc  gtgccatcgc  cgtactcggc  ccggagtgca  ccgagttgct
       112601  ggcggcgctc  gccggcgtcc  cggccgcgac  cgtcgacgag  gccctgttgg
       112651  tgctgcgcag  ggccggcatc  ctggccgccg  accgcgtgga  cttcgtccat
       112701  gacgtcgtcc  gctccgcggt  gctcgacgac  gtcgccccgc  ccaccctggc
   35  112751  cgaactgcgc  accaacgccg  cgctgttgct  gagcgacgcc  ggccgcccct
       112801  ccgaggagct  cgccggccag  ctcatgctgc  tgccggtgct  cgaccagccg
       112851  tggatggccg  cggtgctgcg  cgacgccgcc  gcccaggcgg  agagccgcgg
```

```
      112901  cgccccggag gccggtgtgc gctgcctcta ccgggtgttg gaggtggagc
      112951  cggacaacgt tgccgtccgc atccagatgg cccgcgcgct cgccgagatc
      113001  aacccgcccg aggcgatgcg cctcctcaag gaagcgctct ccctcgccgg
      113051  ggacgtccgc acccgcgccc aggtcgccgt ccagtacggc ttcacctgcc
   5  113101  tcgccgtgca ggaatccccg tccggggtgc ggatgctgga ggacgcgctc
      113151  gccgagctga ccgccgaact gggcccccgaa ccagggcccg tggaccggga
      113201  gttgcggacc ctcgtggagt ccgtgctgct catcgtcggg gccgacgaga
      113251  aggtgacgat cggtgcggtc cgcgaccggg cggcccggct caccatgccg
      113301  cccggcgaca caccggccca gcggcagatg ctggccatga ccaccgtgct
  10  113351  gaccgcgatg gacggccggg acgcccggtc ggccgtcgac caggcccgcc
      113401  gcgccctgcg cgcaccggc gtcgagctgg aaccctggtc gctgctgtcc
      113451  gcctccttcg ccctctccct ggccgacgag gtcgccgacg cgcagtacgc
      113501  actggacctc atgctccagt acggccagga caacgcggcg gtgtggacgt
      113551  acgtcctggc gctctccacg cgcgccctgc tccaccacgg ggtgggcgcc
  15  113601  ttccccgagg ccctggccga cgcccagacc gccgtcgaga tcctcggcga
      113651  ggagcgctgg gcggacggcg ccgtgctgcc ccgtgtcgcg ctggccaccg
      113701  ccctcgtcga ccgcggcgag cccgagcgcg ccgaacacgt cctcgacggc
      113751  atcacccgcc cccgcctcga acgcttcgtc atcgaatacc actggtacct
      113801  ccaggcccgc gcctacgccc gctgggtccg cggggatttc caaggagccc
  20  113851  tggacctcct tctcgcctgc ggtcggtccc tggaggagtc gcgcttcagc
      113901  aacccggcgt tcgtgccgtg gtgggccgat ggcgcggtgc tcctggcgac
      113951  cctggaccgc cacgaccagg cgcgcgaact cgccgcatac ggaagcgagt
      114001  tggccgagcg ttgggggacg gcgcgcggcc tcggactggc cttcatggcc
      114051  cagggcgtcg ccgcacccgg ccgcgccggc atcgatcacc tcaccgaggc
  25  114101  ggtctcgctg ctcgcggact ccccggcccg ggccatggag gcccgggccg
      114151  aacttcttct cggacacgcc cacctgaagc gcgacgacct gcgggccgcc
      114201  cgggaacacc tgcgcgccgc cgccgacctc gcccagcgct gcggcgccgt
      114251  gaagctcggc gtcgacgcca gaaaactgct ggtcaccgcg ggtggtcggg
      114301  tacgcaggat gaccgcctcc ccactcgaca tgctgaccgg gatggaacgc
  30  114351  acggtggcgg acctcgcggt gaccggcgcg agcaaccggg ccatcgcgga
      114401  agccctcttc gtgactgtaa ggaccatcga aacccatctc acgagcgtct
      114451  accggaagct cggggtcggc gggcgtgcgg agctgtccgc cgtcctggag
      114501  accaggaccg ccacctccgg tcggcagccg ccggcctggg tctcccaggc
      114551  acgcggacgc gcttgagacg aacaggacga gaggtagcgg tgccccggag
  35  114601  caaggcgcgg aaccaaccga ccacctgcac gccgcagtgc gcacccgacg
      114651  cccacggcga ccccaccatg ctcctcgaat gcggccggga acagcggctc
      114701  atcggcgacc tcctgcaccg cctcggccag gggcggccat cggtgctcag
```

	114751	cctgaccggc	cggcccgggc	acgcccagaa	cgccctggtc	cgctggggcg
	114801	cgtgccgggc	caggcacgac	gggctgcgcg	tcctgcgagc	ccaggcgacg
	114851	cccgcggaac	gggaactccg	ctacggcgcc	gttctccaac	tgctggccgt
	114901	cctcgacggc	ccgcacggca	gcaccctgga	cgccgcgatc	cgccacgacg
5	114951	gtcccccgcc	actgcccgtg	cccggcatcg	aggaggtgct	gcggcgcacc
	115001	ggcacggcac	ccaccctggt	cgtggtcgaa	gacgtccagt	ggttggaccc
	115051	ggcctcgctg	acgtggttgc	agatcctgct	gcgccacctc	gggccggaca
	115101	ccccgctcgc	cgtcctggcc	agcagctgcg	gtgacaccac	ggccttcgac
	115151	accgacccga	aggcccccgc	cgtcccgggg	ccgccggaca	ccgtgcccgt
10	115201	cgcgcgcttc	gtggtgcccg	cgctcaccga	ccgcggggtc	gccgccaccg
	115251	tccgcgccgt	ctgcggcacc	cccggcgacg	aggagttcat	cgccgcgctc
	115301	acctccgcca	ccgccggcaa	ccccgccatc	ctgcgggacg	ccctgcgcgc
	115351	cttcgtcgac	cacggcctcc	ccgccgacgc	cgaccacctc	ccggagctgc
	115401	acgccctcac	cgctggcgtc	gtcggcgacc	acaccgtgcg	cgccctggac
15	115451	ggcctgcccg	ccgaagtcaa	cgccgtcctg	cgggccctgg	ccgtctgcgg
	115501	cgacctgctc	gacttccacc	gagtccgggc	cctcgccggc	gcgcactcgc
	115551	tgtccgagga	ccggatccgc	accctgctgg	cgagcgtcgg	cctgaccgtg
	115601	tccgtcggcg	acaaggtgca	catccgcttc	cccgcctcca	aggcacgcgt
	115651	catcgaggac	atgcccgccg	cggagcgcgc	cgatctgtac	gtccgcgcgg
20	115701	ccgaactcac	ccacagttgc	ggcgtcaacg	acgaggacgt	cgcccatctg
	115751	ctgctgcgct	cgtcgccgct	cggcgcaccc	tgggtcgtgc	ccctgctccg
	115801	ccgcggattc	gccgccgcgc	tgcgccggga	ggaccaccac	cgggcctgtg
	115851	cctgcctctc	ccgcgccctg	caggaacccc	tcgacccccg	ggaacgcagc
	115901	ctgctgacgt	tggaactggc	cgcggccgaa	gccgtcgccc	ggccggaggc
25	115951	gggggatcga	cgcctggggg	aactcgtccg	cagcaccgtc	gcggacaccg
	116001	accccacgtc	gtccggtgag	ggggtggggg	tccgcgccat	cgacctgggg
	116051	ttcgcccggg	gcaacagcga	atgggtccgc	cgcaccgcgg	gcgaggccct
	116101	gccgtacgcc	gggccggccg	accgggagga	actggtcgcg	ctgttctggt
	116151	tggccgccgt	gcgggacgac	gacgcgccga	tgatccccgt	ggtgccccgg
30	116201	ttgcccgacc	ggccggtgcc	gccggcccag	gccggcgccc	gtgcctggca
	116251	gctggccacg	gcgggggagg	acgcggacaa	ggccaggaag	ctcgcccgga
	116301	tcgccctcac	cggcgggggtg	aacgagagcc	tgatgatgcc	gaaactggcg
	116351	gcctgcgccg	cgctgttcgc	caccgacgac	aacgacgagg	cggtgcacgg
	116401	cctggacacc	atgctcaccg	ccgcccgcag	tgcccacctg	cgcagcatgg
35	116451	ccgcccgcat	tttcaaccta	cgggcccgga	tacacctgtg	cgcggcccgg
	116501	ctggaggccg	ccgaacgcga	tctggacagc	gccgagcgcg	ccctgccgcc
	116551	gacgagttgg	cacccccgtg	cgctgcccaa	cctgatcgcc	acccgcatcc

```
       116601  tcgtcagcat ggagacgggc cgcccggacc gggcccgccg actcgccgag
       116651  gccccggtcc ccgccggcgg cgaggagggt gtgtggtggc ccgccctgct
       116701  gctcgccgc gcccgtgtgg ccgccgacga cggtgactgg gaggaggccc
       116751  tgcggctgtc gcgggagtgc gggcgctggc tttttcgccg gcactgggcc
    5  116801  aacccggcca tgctcagttg gcgccgctg gccgccgagg cgtgtctgaa
       116851  gctcggcgac gtgacggagg cgcgccggct gcgggacgag gagctgttct
       116901  tcgccgaccg ctggggcacc gcgagcgccc gcgggatcgc ccgcctgacg
       116951  acgcggcgac tcttcgacga cgacggcgac cgggccgtcc ggcggatccg
       117001  cgaggccgcc gccctgctcc gcgactcgcc cgcccgcctg gcctacctgt
   10  117051  ggagccggct gagccaggcc ggtgccgaga cggcccacgg cgacaccgcc
       117101  gcggccgcac gctcctggca ggcggtcgcc cggatgaccg ccgcccaccc
       117151  cgccagccgc ctcgccaccg ccgcccgcac cctgaccgtc ccgtccgttc
       117201  cggtcgccac cgcgccgccc accgccgtcg tcccacccgg atggcgcgac
       117251  ctgtccgagg cggagaagga caccgtgctg ctcgccgccc gcggccacgg
   15  117301  caaccgccag atcgccgaac aactcgccgt cagcaggcgc accgtggagc
       117351  tccggctgag caacgcctac cgcaagctga ggatcggcgg acgcaaggag
       117401  ctgtacctgc tcctggaggc gctggaagga ccggtcgcgg atgcttcttg
       117451  agcgggagaa cgaactggcc cggatccggg ccgccctgga cgccgcggaa
       117501  gcgggcgact cctcgctcct cctgatcaac ggtcccctcg gcagcgggcg
   20  117551  ttcggcgctg ctgcgccgga taccggagct ggccggcgac ggcacccgcg
       117601  tcctgcgggc cagcgccgcc tggcgggaac gcgacttccc cttcgggatc
       117651  gcccgccaac tcttcgacca cctgctgtcc ggggcgggcg cgcagggcc
       117701  ggccgaacgc accgccgggg cagagcactt cagccgactg atggacaccg
       117751  gcgaccgccc taccgggacc ggccccgccc tggaggtctc ccaggcagtg
   25  117801  ctccagggcg cccaggcgct gctcgccgac gcgtccgcgg agcggcgcct
       117851  gctgatcctg gtcgacgacc tccagtgggc cgacggcccg tccctgcgct
       117901  ggctggccca cctcacccgg cggctgcacg gcctgcgggc gctgctggtg
       117951  tgcacgctgg ccgacggcga ccaccggggc aggtacccc tggtccggga
       118001  ggtcgccggc gccgcgcaca ccgtcctgcg cctggcgccg ctgtcccggg
   30  118051  acgccacccg cgtcctgctc gccgggcccc agggccggcc gccgcaggac
       118101  gcactggtgc gcgccgtgta cgaggcgtcc aggggcaacc cgctgttcct
       118151  gaccgccttc cggagcgctc tgcgcgccac cggaaggccg cccggcggcg
       118201  accacttcgg cgccgtccgg gagctgagcc cgacggtgct gcgcgatcgg
       118251  ctcgcgggcc atctgcggat ccagccgcag ccggtgcgcg aggtcgcggt
   35  118301  ggcggtggcc gcgctgggcg accacagcga tccggtgctg ctcgcccagc
       118351  tcgccggggt cgatgaaatc ggtttcgccg gtgcccgccg cgcgctggtg
       118401  gacgccggcc tgttggcccg gggacgggac gtccgcttcg tccacggcgt
```

```
        118451   cgtccgcgat gcggtggact ccctgctcac cctcgacgag cgggaacgct
        118501   cgcacgacga cgccgccgat ctgctgtacc gctgcgggcg gccggccgag
        118551   caggtcgccg gccatctgct ggccgtggtc cacccgggcc ggccctggtc
        118601   ggaggcggtc ctgcgctccg cggcccacaa cgcgctgcgc gccggccggc
  5     118651   ccgccgacgc ggcccggtac ctgcgccgcg ccctgctgca ccaccgcacc
        118701   caggacggct gccgcgcccg catcctggtc gatctggcca ccgccgagcg
        118751   cgccctcgac cccgatgcct gtgtacgcca cgtcagccag gcggtcgcgc
        118801   tgctggacac ctcgcgggat cgggccgccg ccgtgttgcg catcccgccg
        118851   tccctgctcg ccgcccccag cccgtccgcc gtcgagttgg tgcggcaggc
  10    118901   cgccgccggg ctcgacgaac cggggcagcg ggacgaggag ggagccgacg
        118951   aactcgccct cgcctggag  gcgtggctgc ggcactccgg ccacgagaac
        119001   cccgtcgagc tggcgtcctc ggtggcgcgg ctgcggcgca tgggggcacg
        119051   gccgccggtg dacagcgtcg ccgaacgcga gttggtcgcc gtgctgttga
        119101   gtgcgggcgc gctcagcggc cggctcagcg ccgcggagat cgccgacacc
  15    119151   ggcaaccgca tcctcgaacg tgagccggcc accgccgccc atgcccacac
        119201   cccgctgccg ctggtgatgc tctcgctgtt cgtcgccgag tccgtgcagg
        119251   gcgtggcctc ctggctggcc agcgaacagc acacccggcg ccggtacgcg
        119301   accggcgcgg acgatgtgct gctgaccgcc gagcgggcct tcgtcctggt
        119351   gacgcagggc cgcccggccg ccgcgcggga gcacgtcgaa cgcgccctgg
  20    119401   tcatggacgc cggcgactgg tcggaacccg ccgtcatgat gttcgccgcg
        119451   gtcgccttcg aactgcgcga cccggccttg agcgaacgca tcctggaacg
        119501   gatccgcgac cgccggccgg ccggactggc gctcaccgcc accggtcaga
        119551   tgctccaggc cgccgtcgac gtgcacttcg gccgggggcg ggacgccctg
        119601   gacacgctgc tggcctgtgg ccgacgcctg gagaccgtgg gatggcgcaa
  25    119651   ttccgcactg ctgccctggc gtccgtatgc aatcgggctg caccagcggc
        119701   tcggcgagac cgatgccgcg ctgcaactcg ccgaggacga gctgaggtgg
        119751   gcccgggagt ggggtgcgac gacgaacctg ggccgggccc tgcgcctgaa
        119801   gggctggctg ttgcaggacg agggactgga tctgctgcgt gagagcgtcg
        119851   agatcctccg cgcttcgtcc tacgcgacgg aactggcgcg caccctcgtc
  30    119901   gtcctcgggc gtcggctgcc gggtgggccg gaggccgagg cggtgttgcg
        119951   ggaggccgcg gggatcgccg ccgcctgtgg tgttccctgg ctcgccgaac
        120001   gggccgaact cggtctgggc agtgccatcg tgccgccggt cgccaccctg
        120051   accccccagcg agcggcgagt ggcgtcgctg gtgagccggg gcctgaccaa
        120101   ccaggccatc gcgaccgaac tcggtgtgag ctcccgggcg gtggagaagc
  35    120151   acctcaccag cgcctatcgc aagctcggcg tctccggccg ccgcgagttg
        120201   gtgaatgcgc tcccgggtcg ttgacgcccg cggccgtccc gttccggacg
        120251   atcacaatcc tctgtgcctg tcagccctgt gggtgcggcg tcgttgtcgt
```

```
        120301  acgcccgtcc ggtcatgaca gcaatcctca acaaaagctc aaacggaagc
        120351  gggggcaggt gcatccaccc catcaccaat gtctcgtatg ttacttttgg
        120401  cccaccgtag ttgggactct gcggaccgta cctcgtcaag tacggaccac
        120451  ggatttaggg gtgcgtagat gtgtcgcgtt gacgtgccga gttgtggcta
    5   120501  actacgctcc gtttcacgga aatcgaaact cttcggcgtg aatcgcggcg
        120551  aacttcgatc gaattctgga ttccgtttct cagaaggcta aaagacgacg
        120601  ggggttatcc gaaaggatgg cggtctcgtg actatcactc acctcactga
        120651  ctcgaatcac cgtacgaccg gcggggtgat atccgctcag accgcaccgg
        120701  ccggggagtc cgtcggcccc gggctgatgg cgtccctcga ccgtgacctc
   10   120751  accatcaagc acgccaacca ggagttccgc cgccgcttcg acgattccgc
        120801  cggtgacgtc tgcggtcgca gcttccggga cctgatgcac ccgagcgtgc
        120851  agcagccgct gatgcgccag ttctcccggc tgatcgaggg caagcggcac
        120901  cgcttcgcct cgcacgtggt cgccgtgggc gcccaggacg cggccttcgc
        120951  gggcaccctc accgcctccg cggtcaccgg caagaccccg gacatcgccg
   15   121001  ggatcctggt gctcatggac tcgtccggcg ccgccgacgc cgccgatgcc
        121051  ggcgtcgtga ccagccagaa gaagttcctg accgagatcg atgcgcgcat
        121101  cctcgaaggc atcgcggccg ggctgtcgac cattccgctc gcctcgcggc
        121151  tctacctcag ccgccagggc gtggagtacc acgtcaccgg gctgctgcgg
        121201  aagttgcggg tgcccaaccg cgccgcgctc gtctcgcgcg cctactccat
   20   121251  gggcatcctg aacgtcggca cctggccccc gaaggtcgtg gacgacttca
        121301  tcaagtgacg cccggccggc ctgtcccgca ccggccggcc accgtcggac
        121351  cgctgtcccg cacgcccgtg tcgccgacgc ccgcctctca cggggcgatg
        121401  cccggcggca cgggcgtttc gtgccctggt cgatgtttat ccattctcac
        121451  gttacgaacc cgcaatgctc atgcgcgcga ttccggccac ttttcgttcc
   25   121501  gtcgtaacca aactgcccgt catatcgttt ccactgaccg tttctgtggt
        121551  ccgcggtgtg ccacctgggc cccactccga cgcgcccgcg acgggtcgcg
        121601  agcggcctcc gaatactcac ggtaaccccc gttcctcctg gtggagttcc
        121651  ccgcgcgtca cattcctccg tgcctcgtgc ggcgtggctc gcccgctttc
        121701  cccgccgccg cgacagccat tgttacaacg gcacgacaga gcggggtgg
   30   121751  tcgacccgtc cgcggtgcgg cgcagtcaag cggaacctct gctcgatgtt
        121801  cgtgccaccc cgaactctgt gttcttggtg accgttttca ccaatgcccc
        121851  gcccgttcgg tggcgaatcc gccgcccgtt cctggccaca acggatcgct
        121901  tcggccgccc accacggccg gcgcagccga ggcttgttgg aatcaccccc
        121951  acccagtcgg ggcgccgttg tggcccggac acgcaggagg aacaccgtgg
   35   122001  atgctgaggg ccgccgcagg gacatgctgg agctgatccg gcgcagcgga
        122051  tccgccgatg tcgtgcgtct cgccgaggag ttcgccgtca gcaaggaaac
        122101  ggtccgccgg gacctgaacg tcctggaggg ccatgggctg atacggcggc
```

```
        122151   ggcacggtgg cgcctatccc atggtgcggc cgggctccga ggccgtcttc
        122201   gtctcccgga ccgcacagcc gatcccggag gagtcccgga tcgccaccgc
        122251   ggccgccgaa ctcctcagcg aggccgagac ggtcttcatc gacgagggct
        122301   tcaccccgca actcatcgcc gacgccctgc cgcgcgaccg gccgctgacc
   5    122351   atcgtcaccg cctcgctccc ggtggtcagc gccttcgcga cgagcccaca
        122401   ggccaacgtg ctgctcctgg gcggccgggt ccgccggggc acgacggcca
        122451   ccgtcgacca ctgggccgtc cacatgctgt ccggcttcgt catcgacctg
        122501   gccttcctcg gcgcggaggg gatctcgcgc aggtacggcc tgaccacccc
        122551   cgacccggcg gtcgccgagg tcaaggccca ggccatccgc gtcgcgcgcc
   10   122601   gcccggtcct cgccggggtg cacaccaagt tcggcacggc gagcttctgc
        122651   cggttcggag aggtgggcga cctggagacg atcgtcaccg gcgccggcct
        122701   gcccgtcgcc gaggcccacc gctaccacct catgggcccc aaggttttac
        122751   gggtgtgacg ccgccggggc gtcccgccgc ccgcccacgg cctggcgcgc
        122801   gccggtgcgg gtcagccgcg tcgcgcgtcc gtcggttccc gccagagtgc
   15   122851   gagcggtgtc acccccaccc ggtagccctc ctcgtcgagg agatggccgc
        122901   cgtgcttcat gtgccagagg gcgaccgcgc gttcggtgac ggtcaggtcc
        122951   gcggcccgtt cggcgagtcg gacctcgaag cgctgggcgt cgtcgaggga
        123001   gcgcacccag gccaccaggt gcaggttgtg ccggctgatg acgctcgcgc
        123051   acaggcggac ctcgcgcatc ccggtgaccc ggcgggtcac ctcgcgcagc
   20   123101   ctcgcggccg ggacctgccc ccagaagctc accgtgaccg gccactcgga
        123151   cagcgggcgg gcgacctcgc accgggcgtg caacatgtcc gcggcgaaga
        123201   gccgttggac gcgacggcgc acggtgtcgg ggccggcgcc gcactgctcg
        123251   gccagcgcgc ggtaggtggc ccggccgtcc acggacagcg cggtgaccag
        123301   ctgttggtcg agctcgtcga ggacgaaggc gggggtgtcc gtgcggtgtc
   25   123351   tggacgcgtc ggcggcgagc cgggcgacct ggtgccggcc gagggcccgc
        123401   aggcgccacc ggctgccctc ggtgtgcacc gggccggcca ggtgcgtccg
        123451   ggcggcccgg acgccgtcca acgcggcgag gtcgtgggtc acccaacggg
        123501   agagcatggc cggatcgcgg gccatgacgt tgagttggag gtcacggtcc
        123551   ccggtgacgt gggagagcgc caccacgtgc ggcgcggcgg ccaacgcccg
   30   123601   cgcgacgtcg agcagtcggc cgggagcgca gtcgatctcg acgaacgcca
        123651   ggcacccctg cccgacgcg gccagcaccg gggccggatg gcagctgatc
        123701   caggcggcgc cggtctcgac cagccggttc cagcgcctgg ccaccgtcac
        123751   ggcgtccagt ccgaggacgg agccgatccg ggtccagctg gcccgcggcg
        123801   tgatctgcag cgcgtgcacc agggcctgat ccacatggtc cagggaccgg
   35   123851   ggggtctgcc cggaatcctg cgccacgggc cacctccttg cgtgtttccg
        123901   gcggattcgg gccgccggtc ggctcaacct tcagcctgga ctcgggtacg
        123951   gccggaccgt accaggcaac ccccggagca acaggagtgg gattcggcat
```

```
      124001  gtcggcagtg  gaacgcgcca  aggaaatgca  acccgagttg  gagcggctgc
      124051  gccggtcgct  gcaccgggag  ccggagttgg  gccttgcgct  gccgcgcacc
      124101  caggagaagg  tgctcgcggc  gctcgacggg  ctgccgttgg  agatcacccc
      124151  gggcaccggc  ctcacctcgg  tgaccgccgt  cctgcgcggc  ggccgccccg
   5  124201  gcggcgcagt  gctgctgcgc  ggcgacatgg  acgcgctgcc  ggtcaccgag
      124251  gagagcggtg  tgccgtacgc  ctcggagatc  cccggccgga  tgcacgcctg
      124301  cggacacgac  ctgcacaccg  ccggactcgt  cggcgcggcc  aggctgttgg
      124351  ccgaacgccg  ggaacgcctg  cacggcgacg  tggtgttcat  gttccagccc
      124401  ggcgaggagg  gccacaacgg  ggccggcgcg  atgatcgagg  agggcgtgct
  10  124451  ggacgcggcg  ggcaagccgc  tcgacgcggc  gtacgccctg  cacgtggcgt
      124501  ccaaccaggt  gcccggcggc  gtggtcatca  cccgctccgg  caccatcacc
      124551  tcggcctccg  acctgctgac  ggtgacggtg  cggggcgagg  gcggccacgg
      124601  ctcgacgccc  gccacggcca  aggacccggt  cccggcggcc  tgcgagatgg
      124651  tcaccgcggt  gcagaactgg  gtgacccgcg  ccttcgacat  cttcgacccg
  15  124701  gtcgtggtca  ccgtcggcac  cttccacgcc  gggacgaagg  cgagtgtcat
      124751  ccccgacaac  cgccgagttc  caggccacgg  tgcgcagtta  ttccgaatcc
      124801  gcccgggacc  ggctcgccga  agggaactgc  cccgtctggt  gcgggccatc
      124851  gcggccgccc  acggcctgga  agcggaggtc  gactacctcc  ggcactatcc
      124901  ggtcaccgtc  aacgacacgg  acgagaccgc  cttcgcggtg  gccaccgccc
  20  124951  gcgacctgtt  cgcgcccggc  gaggtgtggg  agtcgccgat  acccaccaac
      125001  ggctccgagg  acttcgcgta  cgtcctccgg  cgcgtccccg  gcgcgatgct
      125051  gctggtgggc  gccgccccg   aggggagcga  ctggcagcac  gcgccgatga
      125101  accactcgcc  gcgcgccgtc  ttcgacgacc  gggtgctgta  ccggcaggcg
      125151  gcgctgctga  ccgaactcgc  cgcgcgccgg  ctcgcggcca  ccgcgccggc
  25  125201  cgagcccgcc  gtggcgggat  gacaccgccg  ccgcacgcgt  ggggcccgca
      125251  cgcgtgcggc  gtacgggccc  tggccggtct  ccttcccggc  tacttgcgcg
      125301  cggggcactc  cttccacgcg  aagtggtaga  gggtgttgat  gctgccgtcg
      125351  gtggagtcat  ggccatgtag  ctggtggcct  tggggtcagc  cccgctgccg
      125401  t
```

SEQ ID NO. 36

NysDII

```
    1  msftypvsmp  wlqgreldyv  teavgggwis  sqgpyvrrfe  eafaayndvp
   51  fgvacssgtt  altlalralg  vgpgdevivp  eftmiasawa  vtytgatpvf
  101  vdcgddlnid  vsrieekitp  rtkvimpvhi  ygrqcdmdav  lnlayeynlr
```

```
    151  vvedsaeahg  vrprgdiacf  slfankiisa  geggvclthd  phlaeqmahl
    201  ramaftkdhs  flhkklaynf  rmtnmqaava  laqteqldti  lalrrdiekr
    251  ydealrdipg  itlmpprdvl  wmydlraerr  delcaylage  gietrvffkp
    301  msrqpgyfsa  dwpalnaarl  sadgfylpth  tgltaqeqef  itgrirafyg
    351  va
```

SEQ ID NO. 37

NysI

```
      1  mdneqklrdy  lklatadlrr  trrrvhkles  aaqepvaiig  mtcrypggvr
     51  spedlwrmve  agehgvtpfp  tdrgwdleal  aaaptasggf  lhdapdfdad
    101  ffgispreav  amdpqqrvvl  esaweafera  gidptsvkgs  rtgvfigama
    151  qdyrvgpadg  aegfqltgnt  gsvlsgrisy  tfgtvgpavt  vdtacssslv
    201  avhlatqalr  agectlalag  gvtimsgpgt  fiemgrqggl  sadgrcrsfg
    251  dtadgtgwae  gvgilvlerl  sdavrnghei  lavvrgtavn  qdgasnglta
    301  pngpsqqqvi  qqalvnarla  agdidvveah  gtgttlgdpv  eaqallatyg
    351  qnrpadrpll  lgsvksnlsh  tqaaagvagv  ikmvmamrhg  tlprtlhaee
    401  pthhvdwsqg  avrlltdttd  wpatgaprra  avssfgisgt  nahtiieqap
    451  epqpedaata  qddaagstpa  tapvvpgvvp  vllsgrtpda  lrgqaaalra
    501  aldtgrrpdl  ldlahslatt  ragfehravl  latdhpaltd  gltaladadd
    551  paaapawitg  ttraetrlav  lftgqgaqrl  gagrelaarf  pafataldaa
    601  ldaftphldr  plrevlwgtd  aalldrtaya  qpalfaveva  lyrliesfgv
    651  rpdhlaghsv  geivaahlag  vlsladaatl  vaargrlmqa  lpdggamiav
    701  qaseadvapl  laghedqvai  aavngpsavv  lsgaeatvta  laeqlaadgr
    751  ktrrlrvsha  fhsplmepml  dafravvedl  tlqppllpvv  snltgkpatv
    801  aqltsadywv  dhvrhavrfa  dgidwlarhd  ttaflelgpd  gvlsamaqdc
    851  ldaadadavt  lpalragrpe  ehtlttalag  lhvhgatldw  tgcfagtgar
    901  rtdlptyafq  rrrywpkalq  sgtadlrsvg  lgaahhplls  aavsladagg
    951  tlltgrlsrq  thpwladhtv  rgttllpgta  flelavragd  evgcdrveel
   1001  tlaaplllpe  qggvqvqlwi  gnpdvsgrrt  vnvharpdtg  ddtpwtahat
   1051  gvlttadasr  qlpasseqgg  tplagdphpa  ldaaqwppag  aeplpldghy
   1101  drladggfgy  gpvfqglraa  wrggdvvyae  velpeagrsd  aeafglhpal
   1151  ldaalhaapf  tglgergrgg  lpfswegvsl  haggattlrv  rltpvaddal
   1201  altvadgtga  pvlsvdslvl  rsvatqqldt  aaavardalf  rldwtpvqpt
   1251  atdpgpvall  gadpfgllth  agfadapayp  dlaalaaadg  pvpttvvlsl
   1301  agtgddaadp  arsahrcaae  alaavqtwld  hherfaaarl  vfvtrgatvg
```

```
      1351  rdvaaaavwg  lvrsaqsenp  gcfalvdldp  dgavgaaalv  aalvsgepql
      1401  avrgdvlrva  rlvrrpltev  gagadgtgdg  vgdgsgvsfs  gegavlvtgg
      1451  tgglgavlar  hlvaeygvrd  lllvsrsger  avgagelvae  lagvgarvrv
      1501  vacdvtdraa  vvelvgghav  savvhaagvl  ddgmvgaltg  erlsavlrpk
   5  1551  vdavwhlhea  trgldldafv  vfsslagvfg  spgqanyaaa  nafldalmtr
      1601  rraeglpgls  lawgpweqsg  gmtgtltdvd  aerlarsgvp  plsvaqglal
      1651  fdaavagtda  tcvpvrldlp  vlrargevpp  llrslirvra  rraavagsat
      1701  agnlaqrlrr  ldedgrdemv  ldlvrgqval  vlghatggdv  dagrafrdlg
      1751  fdsltavelr  nrlntvtglr  lpatlvfdyp  tvrhlatyvl  dellgtdaev
  10  1801  atvqpaavav  addpivivgm  acrypggvss  pedlwrvlte  gtdavsgfpt
      1851  nrgwdvesly  hpdpdhpgts  ytrsggflhe  agefdpgffg  msprealatd
      1901  sqqrllless  weaieragid  pvslrgsrtg  vfagvmysdy  samlaspefe
      1951  gfqgsgssps  lasgrvaytl  glegpavtvd  tacssslvam  hwamqalrsg
      2001  ecglalaggv  tvmstpavfv  dfarqrglsp  dgrckafada  adgvgwsegv
  15  2051  gvlvlerqsd  avrngheila  vvrgsavnqd  gasngltapn  gpsqqrvirq
      2101  alasggltag  dvdvveahgt  gttlgdpiea  qallatygrd  reperplllg
      2151  svksnlghtq  aaagvagvik  mvlamrhgvv  prtlhvdaps  shvdwsegav
      2201  ellseqaawp  etgrvrragv  ssfgisgtnv  hviveqapga  kaiaaagaar
      2251  rtpgavpvll  sgrgrsalrg  qaarllghlq  arpdaelvdv  alslattrsr
  20  2301  feqraavvaq  drdqliaslg  alaadrpdpa  vvegeaagrg  rtavlftgqg
      2351  sqraamgrel  hevqpefaaa  fdavcavfdp  lldrplrevv  faedgsdeaa
      2401  lldetgwtqp  alfavevalf  rlveswgvrp  dfvaghsige  iaaahvagvl
      2451  tledacrlva  aratlmqalp  tggamiaiqa  tedeiaahld  dtvaiaavng
      2501  pqsvvisgde  eaaetiaatf  aergrktkrl  rvshafhspr  mdgmldafri
  25  2551  vaegltyrap  riplvsdltg  rraddaevct  aeywvrhvre  avrfadcvrt
      2601  lrdagattfl  elgsdgllta  maedtlgddh  daelvpmlra  graeelaaat
      2651  alarlqvrgv  dvdwaaylag  tgarrtdlpt  yafqhayywp  qlptpaaala
      2701  aadpadqqlw  aavergdare  ladilglgeq  dltpldsllp  altswrrgnq
      2751  ekhlldtlry  rvewtrlskp  tapvldgtwl  lvasdataad  qpalldglad
  30  2801  algshgarvr  rlllddscad  ravlaerlar  tadvdaatqv  lsvlplderd
      2851  addcppltrg  laltvalvqa  ladtgaqgrl  wtatrgavst  npadpvthpv
      2901  qaaawglgrg  valehprlwg  glvdlpqvfd  eragqrlagi  lavkdapdge
      2951  dqvalratgv  sgrrlvrhtv  ealptaaeft  atgtvlitgg  tgglgaevar
      3001  wlaragaqhl  vltsrrgpda  pgaaelrael  egygpsvsvv  acdvadrdal
  35  3051  aavltalpee  lpltgvvhta  gvghygpldt  lstaefaglt  aaklagaahl
      3101  dalladreld  ffvlfgsiag  vwgsgnqsay  gaanayldal  alhrrargla
      3151  atsvawgpwa  eagmaaddav  setlrrqglg  lldpapamte  lrravvrqdv
```

```
       3201  tvtvadvdwq  ryaplftsar  psaliaglpe  vralaadert  eqdatgasev
       3251  vtrvralaep  eqlrlltdlv  rtesatvlgh  ssadavpegr  afrdvgfdsl
       3301  tavelrkrlg  aatglslpst  mvfdyptple  laqylraeil  gavlevagpv
       3351  atggaddepi  aiigmacrfp  ggvsspeqlw  dlvasgtdai  sefpvnrgwq
   5   3401  tghlfdpdpd  rpgttystqg  gflheadefd  ptffgispre  alvmdpqqrl
       3451  llettwesfe  ragirpetlr  stltgtfvgs  syqeyglgag  dgteghmvtg
       3501  sspsvlsgrl  syvfglegpa  vtvdtacsss  lvalhlacqs  lrngesnlav
       3551  aggatimttp  npfiafsrqr  alakdgrcka  fsddadgmtl  aegvgvvlve
       3601  rlsdaqrngh  pvlavlrgsa  inqdgasngl  tapngpsqqr  virqalanar
  10   3651  lapgdidale  ahgtgtplgd  pieaqalfat  ygrdrdpesa  lllgsvksni
       3701  ghtqsaagia  svikmvmalr  hselpptlha  dapsshvdws  agtvrlltqa
       3751  rawpetgrpr  raavssfgis  gtnahvlleq  apvadtpaee  rpavapvpia
       3801  agvvpwvvta  rsaaalrgqa  erllahaetv  gtalpaagpl  diglslvsar
       3851  arfehravvv  ppagtdplaa  lravatdgps  pvvargvadv  egrtvfvfpg
  15   3901  qgsqwvgmgs  qlldesavfa  eriaecaaal  aeftdwslvd  vlrgvvgaps
       3951  lervdvvqpa  sfavmvslaa  lwrsrgvlpd  avvghsqgei  aaavvsgals
       4001  lrdgarvval  rsqaigrala  grggmmsval  svdvleprlv  efegrvsvaa
       4051  vngprsvvva  gepealdalh  arltaddira  rriavdyash  shqvedlhee
       4101  llevlaelap  rtsevpffst  vtgdwldtar  mdagywfrnl  rgrvrfadav
  20   4151  adllaaeyra  fvevsshpvl  smavqeaide  agvpavaagt  lrrdqggtdr
       4201  fllsaaevfv  rgvdvdwagl  fegtgasrid  lptyafqheh  lwavppapea
       4251  vaaadpddaa  fwtavedgdv  saltaalgtd  edsvaavlpa  ltswrrarrd
       4301  rstvdawryr  vawkplggtl  phpsltgtwl  lvtadgiddt  dvagaletyg
       4351  aevrrlvlde  ecvdravlre  rlagaedvtg  ivsvlaaaer  tdavpgtslv
  25   4401  lgtaltvali  qalgdaeida  pvwaltrgav  stgradelta  pvqaqvtgig
       4451  wtaalehpqr  wggtldlpaa  ldaraaqrla  avlsgalgsd  dqlairpsgv
       4501  ftrrivraea  tagrpagtwt  prgttlvtgg  sgtlaphlar  wlaqrgaehl
       4551  vlisrrgtaa  pgaaelvael  aesgteatva  acditdrdav  aalladlkad
       4601  grtvrtvvht  aatielhtld  attladfdrv  lhakvtgaqv  laellddeel
  30   4651  ddfvlyssta  gmwgsgahaa  yvagnaylaa  laehrrangl  palslswgiw
       4701  addlklgrvd  pqmirrsgle  fmdpqlalsg  lqralddnen  vlavadvdwe
       4751  tyhpvytsgr  ptplfdevpe  vrrltaaaeq  sagtvaegef  aaalralsda
       4801  eqqrtlletv  rteaasvlgl  ssaedltdqr  afrdvgfdsl  tavglrnrla
       4851  svtgltlpst  mvfdypnpaa  laaylhgela  garsaaagaa  avptgapdad
  35   4901  dpiaivgmsc  rypggvgsae  dlwrialdev  daisgfpadr  gwdaeglydp
       4951  dpdrpgrtys  vqggflrdva  efdpgffgis  prealsmdpq  qrllletawe
       5001  afehagidpv  gqrgsrtgtf  vgasyqdyas  gvpnsegseg  hmitgtlssv
```

```
5051   lsgrvsylfg  fegpavtldt  acssslvamh  lacqslrnge  sslalaggvs
5101   imstpmsfvg  fsrqralaed  grckayadga  dgmtlaegvg  lvllerlsda
5151   ranghqvlav  irgsavnqdg  asngltapng  psqqrvirqa  lansavapgd
5201   idvleghgtg  talgdpieaq  allatygqdr  aperplllgs  vksnightqm
5251   asgvasvikl  vralqegvvp  kslhidrpst  hvdwssgaig  lltertpwpe
5301   tgrprraavs  sfgisgtnvh  tileqapade  aptpadpprd  glvpvllsgr
5351   geaalraqaa  rllafveerp  eahltdlahs  latsraaler  raaviaadrd
5401   tltrglrals  dgrpdpglvq  gtagrgrtaf  lftgqgsqrp  gmgrelhdry
5451   pvfadaldev  larlddgpdr  plrevlfaap  dsaeaalldr  tgyaqpalfa
5501   vevalfrllt  swgltpdyla  ghsvgelaaa  hvagvlsldd  actlvaargr
5551   lmqalpegga  mvaleaaede  vlpllegltd  rvsvaavngp  rsvvvagvee
5601   dvllladlfa  adgrrtkrlr  vshafhsplm  damlddfaav  argltyhppt
5651   ipfvsnvsgg  lataeqvrtp  dywvghvraa  vrfadgidwl  atqgdvhtfl
5701   elgpdgvlsa  maresltdps  rtallptlrg  drpeepalvt  avaaahahga
5751   rvdwsgyfad  hgarrttlpt  yafqrerywp  dttaatsaht  pgsaldaefw
5801   aaverddvaa  laasldldda  tvtamvpalt  awrrrrgeqt  eldswryrvt
5851   wkprggatap  aaltgrwlvl  vphdhqdrqd  dataawaadv  etalgtttvr
5901   ltvtttdraa  laariteaag  dqgpfsgvls  llplatgdag  hpgapaaltl
5951   tttavqalgd  agidaplwnv  trgavavgra  eqvtapeqaa  vwglgraval
6001   elparfggtl  dlpatldgqa  arrlravlaa  tdgedavalr  psgvflrrla
6051   hapagpdtar  tafdpaagtv  litggtggig  ghvarrlard  gathllltsr
6101   rgpaapgada  lraeleelga  rvtlaacdaa  drdalaalla  elpddaplca
6151   vfhtagvved  hvvdaltpen  faavlraktv  aahhlhelta  dldlaafvlf
6201   sstagvlgaa  gqgnyaaana  hldalaehrr  shgltalsva  wgpwagsgmv
6251   adaaeltdrv  rrggfeplap  epavrallra  ienddttval  adidwerfqr
6301   afaavrplpf  vadlpetgra  tpatatgaat  glrqqlaelp  eherpaavld
6351   llrtqvaavl  ghadprtved  dhafrdlgfd  sltilelrna  lnaatglslp
6401   atlvydlptp  remadfllae  llgtlptdta  atvastaspk  lsasfeqggt
6451   pfddpiavig  igcrfpggvt  tpeelwqlld  egrdgisrfp  ddrgwdlaal
6501   gagasdtleg  gfltgvadfd  arffgispre  alamdpqqrl  llettweale
6551   ragidpttlr  gsttgvfvgt  ngqdyptllr  rsasdvagyv  atgntasvms
6601   grlsyalgle  gpavtidtac  ssslvalhwa  gralragecd  lvvaggvsvm
6651   aspdsfvefs  tqgglapdgr  ckafsdaadg  tawsegvgil  vlerlsaarr
6701   nghqvlglir  gtavnqdgas  ngltapngls  qqrviaqala  darlrpadid
6751   aieahgtgtt  lgdpiearal  itaygrdrda  erplllgtvk  snightqaaa
6801   gaagvikmlm  amrhgtlprt  lhvgtpsshv  dwsggtvall  ddarpwprtg
6851   qprragvsaf  gvsgtnahvv  veqapeteap  aapaaepape  atptvvpwvv
```

```
        6901   sgrsrealqa  qldrltahta  ahparsaadv  grslatdrtl  fphravllag
        6951   pdgvreaara  aaprtpgrta  flfsgqgaqh  almghdlyqr  fpvyadaldt
        7001   vlaqfdtvld  vplraalfaa  pgtpeaalld  qtgftqpalf  avevalfrla
        7051   eswrltpdfv  aghsigeiaa  ahvagvfsle  dactlvaara  slmqqlprdg
    5   7101   amvaleated  evaplltdgv  alaavngprs  vvvagaedav  ravadrlaad
        7151   grrtrrltvs  hafhsplmdp  mltdfarvae  gltyheprip  lvstllgapa
        7201   gaelrtpdyw  vrhvretvrf  adgvralhda  gagtfveigp  dgvltaltqq
        7251   tldtveagap  avvvplqrrd  ragdlalleg  latlhthgtg  pswpayfeat
        7301   gghrtdlpty  afqrerywpe  lgapvatapq  dpaawryhet  waplpapeaa
   10   7351   apagralvlv  pagnrdtawm  tavadalgad  tvtaepdala  eqltaagdtp
        7401   wrvvvsllaa  aseglpadga  wpaallatld  eagvhaplwc  vtrgavavag
        7451   eaptavgqaa  lwglgrvaal  dhpdrfggla  dlpadtdaha  agllaahlaa
        7501   pgteaeiavr  atgvharrlv  rtpaaadgat  wlptgtvlvv  ggtggtgtmg
        7551   graarwlvre  garhlvltap  dgtttaadte  altaelaalg  aritvvdhdp
   15   7601   tapdgfaall  dglpddtplt  avvyapeada  apgtaaelsa  alapvtalga
        7651   altgrpldaf  vlfgsiaglw  gvrgraaeaa  sgayldafar  acrdrgtpal
        7701   avawgawadl  vgpslaahlr  mnglpvmdad  taltalsrav  adgsaaeava
        7751   dvrwetfapl  hhearrtalf  dalpeargal  aeaardradr  ktaagdygrw
        7801   laeqpaadhd  aillalvtek  aatvlghadh  dllepdlpfr  dlgfdsltav
   20   7851   dlrnqltaet  gltlpatlvf  dhpnpaalaa  hlraqllgea  sdsaapvaap
        7901   valgadddai  vivgmacryp  ggvtspedlw  qlvgdevdav  gdfptdrgwd
        7951   laalagdgpg  rsataqggfl  ydatdfdpgl  fgisprealv  mdpqqrille
        8001   tswealerag  idpatlrgsg  ttgvfvgggs  gdyrppeeag  qwqtaqsasl
        8051   lsgrlaytfg  iqgptvsvdt  acssslvalh  laaqalrage  csialaggvt
   25   8101   vmatpvgfve  fsaqgalspd  grcrafsdda  ngtgwsegvg  mlvverlsda
        8151   rrnghrvlav  lrgsainqdg  asngltapsg  paqqrvirqa  lanarlrpad
        8201   idaveahgtg  trlgdpieaq  allatygqdr  erpvllgslk  snightqaas
        8251   gvggvikmvl  amqhgelprs  lyaenpsshv  dwtagrahll  tartpwpdsg
        8301   rprraavssf  gasgtnahai  leqppreelp  arpaddgapl  pfllsgrsqn
   30   8351   alraqarrll  arltahpdtr  aadlayslat  traafehraa  itatdhdglr
        8401   tgltavaegt  taphtaehhl  qgtgkravlf  sgqgsqrlgm  grelherhpv
        8451   faeafdsvla  rlddrldtpl  rdvvwgtdee  alhatgntqp  alfavevaly
        8501   rlieswgvrp  dfvaghsvge  laaahvagvl  slddacrlva  araalmqrlp
        8551   aggamiavea  tedevtpllt  dgvslaavng  ptavvlsgag  davtalgqal
   35   8601   aerghrttrl  rvshafhshl  mdpmladfrt  vaegleyhpp  ripvvsnltg
        8651   dvadaadlcs  adywvrhvrg  tvrfadgvrt  madrgvhlfl  elgpdavlsa
        8701   marqcapdav  vvpalrrnrd  edetlvgava  rlhvhgagpr  wdayfagrga
```

```
         8751  qwldlptypf  qrgrfwpesl  pgaasaapaa  gqpaetdaaf  wdavaqedft
         8801  alesvldves  dalskvlpal  mdwrsrqade  sqlagwrhri  vwkrltgaal
         8851  ahrkalsgtw  lavvpegfad  dpwvtttldg  lgthlvhlev  aeadraalad
         8901  aiaartadgt  rfggvislla  lreeltgavp  egtalttttll  qalgdagvda
    5    8951  plwcvtrsav  sagrtdrphr  plqgavwglg  rvaaleypqr  wgglvdlpee
         9001  pdersaagla  avlagldged  qvavrgtavl  arrlvpapgr  kpsrpwhpsg
         9051  tvlvtggtga  lgahvarrla  kdgaqhlvll  srrgpdapga  aelraeldal
         9101  gtdvtvaacd  vadrdqltav  ldalpadrpl  tgvvhtagvl  ddgvldrltp
         9151  erfqevfrak  vtsallldel  trdrelaafv  lfssasaavg  npgqanyaaa
    10   9201  navldalaeq  rrvlglpats  vswgawgggg  madadgadea  arragvgamd
         9251  phlaveallr  lvaekeptav  vaevaldrfa  gafggsrpsa  llrefpgyre
         9301  alaaqaeqaa  dggglaarla  alpparrldt  vvdlvrtraa  qvlgypdtea
         9351  vaaersfrdl  gvdslgavel  rnqlsaatgl  nlpatlvfdh  ptplvlgehi
         9401  lgglfpdepa  gsddeteira  llasvpldql  reigvlepll  qlagrggraa
    15   9451  dgddgesvds  mtvadlvraa  lngqsdl
```

SEQ ID NO. 38

NysJ

```
            1  mnapenpetp  ennvvaalra  avketdrlrr  qnrmlvaaak  epiavvgmac
           51  rfpgavdspe  alwemvatgt  dvisgfpddr  gwdlealrns  gtdardtdvs
          101  qrggfldcia  dfdpgffgis  preavtmdpq  qrlllttawe  averagidat
          151  tlratrtgaf  igtngqdyay  llvrslddat  gdvgtgiaas  aasgrlsytl
    25    201  glegpaltvd  tacssslval  hlavqalrng  ecgmalaggv  nvmatpgslv
          251  efsrqgglar  dgrckafada  adgtgwsega  gvlllerlsd  aqrnghpvla
          301  vvrgsavnqd  gasngftapn  gpsqqrvirq  alanaglatg  didaveahgt
          351  gtplgdpiea  qsilatygqd  rahpvllgsi  ksnmghtqaa  sgvagvikmi
          401  mamrhgvlpr  tlhvdrpsth  vdwttgsvel  ltdahpwpet  grprrtgiss
    30    451  fgvsgtnahv  iveqapdtpa  eaaddtpprt  prtlpwllsa  rtgaalrdqa
          501  talldhldrp  dgdrgptald  tafslattra  alehrlavvt  gtdgtagrda
          551  ltawlahgta  pdaheghaag  rtrcaalfsg  qgaqrlgmgr  elharfpvfa
          601  raldtavdll  daelggtlre  viwgtddapl  netgftqpal  favevalyrl
          651  ieswgvapdf  vaghsigeia  aahvagvfsl  edactlvaar  aglmqalprg
    35    701  gamvaveate  devsplltdg  vaiaaingpt  slvvsgdeta  tlavaarlae
          751  qgrrttrlrv  shafhsplmd  pmlaefrava  eglsygepqi  pvvsnltgav
          801  adgtllgtad  ywvrhvreav  rfadgiralt  dagvgaflel  gpdgtlaala
```

```
       851  qqsapdavsv  pvlrkdrdee  paavaalarl  htagvpvdwt  afyagtgahr
       901  tdlptyafqy  erywpkatyr  padatglglt  aadhpllgaa  msvagsdell
       951  ltgtlslath  pwladhvvgg  mvffpgtgfl  elavraadqv  gcdrveelml
      1001  aaplilpatg  tvqmqiavga  adddggrdlr  fftrpgddpd  aawaqhatgr
 5    1051  itegervlal  dtttwpprda  epvdidglyd  ryrangldyg  pvfrglravw
      1101  rrdteiyaev  alpegtadad  afglhpalfd  avlhstlfas  adgddrsllp
      1151  fawngvslha  agadalrvri  tscgpdavei  tavdpqgrpv  vsvesltlra
      1201  agpdagtadh  radagslfrm  dwtprtvhap  atpatwavlg  tdpiglteal
      1251  taagpdtvtg  lrdgvdalge  ltagddrpvp  dvvavplrga  tdhgpagahd
10    1301  ltrtvlallq  ewlaeerfar  srlllvtrga  vadgergpld  laaapvwglv
      1351  rsaqsenpgr  lllvdlddta  esaaqlpllp  alldadepqa  vvregtvrvg
      1401  rlarldsgrg  lvpppgtpwr  lgsrakgsld  glallphpea  rrpltghevr
      1451  vgiraaglnf  rdvlnalgmy  pgdaglfgse  aagvvvevgp  evtglapgdr
      1501  vmgmlfggfg  plgiadarll  tpvpadwswe  tgasvplvfl  tayyalkelg
15    1551  glragekvlv  hagaggvgma  aiqiarhvga  evfatasegk  wdvlrslgva
      1601  ddhiassrtl  dfeaafaeva  gdrgldvvln  alsgefvdas  mrllgdggrf
      1651  lemgktdira  adsvpdglsy  hsfdlgmvdp  ehiqrmlldl  velfdrgala
      1701  alpvrswdvr  rageafrfms  laqhigkivl  tvpqpldpdg  tvlltggtgg
      1751  lagllarhlv  tehgarhlll  agrrgpdapg  aaalhaelta  lgaevtvaac
20    1801  dvadrtalaa  llatvpaehp  ltavvhtagv  lddgtltaln  pdrlatvlrp
      1851  kvdaawhlhd  ltrhldlaaf  vlysstagvm  ggpgqanyaa  gntfldalaa
      1901  hrhalglpat  slawgaweqg  agmtgaltdh  dlrrvsdagg  qplltaergl
      1951  alydaataad  eplivplglt  ggalpagvgv  pavlrglvrt  agrraragta
      2001  gvsraglaer  laalpeeert  pflvelvrte  aatvlghgst  dpvdarrefr
25    2051  qlgfdsltai  elrnrlgkat  gltlpatlif  dyptpdrlav  hlhdellgad
      2101  apvtvtaaaq  aadpehdpvv  ivgmscrfpg  gvsspeelwd  lvasgtdait
      2151  gfpadrawdr  hpqlagapga  rtgqggflrd  iadfdaaffg  isprealamd
      2201  pqqrilleva  weaaeragid  pqtlrgsdtg  vfmgvsgqdy  aglvmrsrdd
      2251  iaghattgla  vsvvsgrlay  alglegpals  vdtacssslv  slhlaaqalr
30    2301  agectmalag  gvtvmttaan  ftgfsrmggl  aqdgrckafs  dsadgtgwse
      2351  gaavlvlerl  sdarraghrv  lavvrgsavn  qdgasnglta  pngpaqqrvi
      2401  rqalanaglt  pvdvdaveah  gtgtplgdpi  eaqaliaayg  tdrdpehpll
      2451  lgsvksnigh  tqsaagaagl  vkmvmamrhg  ilpqtlhlte  psshvdwsag
      2501  tvrllterta  wprtdrprra  gvssfgisgt  nahvileqpp  aeptpaadpg
35    2551  rpaptvvawp  vsaqtpaald  aqldrlrtaa  alapldtaht  latgrslfeh
      2601  ravllatvgd  patgapdlpe  vargaatphr  taflfsgqga  qrsgmgrelh
      2651  aafpvfaaaf  devvavldae  lgsdadggvs  lrevmwgggs  elldrtrftq
```

```
      2701  palfaveval  frlvaswgvg  pefvaghsvg  eiaaahvagv  fslvdacrlv
      2751  varaslmdal  pvggvmvave  aaeaevvpll  vdgvaiaavn  gpvsvvvsgv
      2801  eaavgqvvdq  lvergrrvrr  lavshafhsp  lmdpmldafr  avaegleyhq
      2851  pripvvsnvt  gevaaaeelc  aadywvrhvr  atvrfadgvr  tlaergataf
  5   2901  leigpdgvls  alargvlpae  alvtptlrkd  rdeesallag  larlhvagvt
      2951  vdwsaaltgt  gargtdlpty  afqrerywpe  laaepaggga  daadaefwaa
      3001  veradatala  ahldidgdql  gavlpalsaw  rtrrrttsat  nalrhreswe
      3051  plslagtpht  ggvlvlvpaa  attdpwvadv  vaalgpdarr  vdvpadgtdr
      3101  aalaalltea  addtaptavv  sllaldetsg  ddavpagtta  taalvqalad
  10  3151  tgapaplwal  trgavaalpd  eqptapaqaa  vwglgriaal  elprhwgglv
      3201  dlpadldert  arrlpaalad  agdedqlalr  atgaygrrit  papapddapg
      3251  tgwqptgtvl  itggtgalgr  htarwlaahg  aehllllsrs  gpdapgaael
      3301  tteltalgar  vtlvacdaad  reqltrvlae  vprdcpltgv  vhtagvlddg
      3351  vltgltpdrf  atvfrakvas  avlldeltrd  tdlavfalfs  svagavgnpg
  15  3401  qagyaaanav  ldalaarrra  qglagtsiaw  gawagdgmaa  rhtrpgaepv
      3451  glldpdlavp  alaravtepq  ptlvladlqq  prllesllal  rpspllsrlp
      3501  aartaaravq  eadrrragaa  adlrdqlagt  apadrhavll  rlvrttaaav
      3551  lghtgadair  adkpfrdlgf  dsltavelss  alaaatglal  ppslvfdhps
      3601  praladhlra  eltgdrpesa  paappapvpa  adddpivvvg  macrfpggvt
  20  3651  tpeefwqlla  egrdgidafp  tdrgwdldvl  grrrpgpqrp  prsaassyda
      3701  aafdpgffdi  sprealamdp  qqrllletaw  eavertgtdp  trlrgsrtgv
      3751  fvgtngqdya  glvlraqedv  eghagtglaa  svisgrlaya  fgfegpavtv
      3801  dtacssslva  lhwavqalra  gecslalagg  vtvmttstsf  agftrqggla
      3851  pdghckafsd  sadgtgwseg  vgvlvverrs  dalrngheil  avvrgsavnq
  25  3901  dgasngltap  ngpaqqrvir  qalanaglap  gdvdaveahg  tgtvlgdpie
      3951  aqallatygq  drpadrplwl  gsvksnight  qaaagaaglm  kmvlalqhgt
      4001  lprtlhvtep  strvdwsaga  vrlltertvw  prtdrprrag  vssfgisgtn
      4051  ahvileqppa  eptptapadr  ptrtpavlpw  vvsarsatal  daqlarlraf
      4101  aaerpdlppa  dvahslvtsr  atfehravll  aapdgitaaa  raearersta
  30  4151  flfsgqgaqr  sgmgrelhaa  fpvfaaafde  vvavldaela  tgsgggvslr
      4201  evmwgggsel  ldrtrftqpa  lfavevalfr  lvaswgvgpe  fvaghsvgei
      4251  aaayvagvfs  lvdacrlvva  raslmdalpv  ggvmvaveaa  eaevvpllvd
      4301  gvaiaavngp  vsvvvsgvea  avgqvvdqlv  ergrrvrrla  vshafhsplm
      4351  dpmldafrav  aegleyhqpr  ipvvsnvtge  vaaaeelcaa  dywvrhvrat
  35  4401  vrfadgvrtl  aergatafle  igpdgvlsal  aaaclfdtda  evvpalrkgr
      4451  peehtaltaa  aqlhvagvdi  dwtavlagtg  grrialptya  fqrerywpsl
      4501  aaqapgdagg  lgleagrhpl  lgaattvags  aeilltgrls  ttaqpwlavy
```

```
       4551  eadgrtvlpa  avlaelavra  gdqadcptva  eltvaaplvl  tgaaaqrlqv
       4601  rvaapddtgr  ralsvharpd  dspdspwtlh  atavlthdtp  qppapdtgwp
       4651  peravpldal  ptatgparia  aawqwgdelc  aeielpepgp  aerafalhpa
       4701  lldtavragg  lldgdatlda  lgwrglalha  asatalrvrl  tpdgtdtwal
   5   4751  eatdpqgapv  vsvtgltlgt  ptvdrsgaga  addgatlldl  ewvpapqaap
       4801  tggdhlpyav  lgdqlaeldg  qlriagdgpg  rvaslaalld  ggaplprlvl
       4851  apvlgvptge  gdlpaavrgt  ttavlellqr  wtadartads  hlvivtrgav
       4901  aagaedvhdl  aaapvwglvr  saqsehpgsf  llldldpadp  agasraaapa
       4951  tlaalldage  tqaavradtl  tvarltraad  gpeataghpv  rdwdrdgtvl
  10   5001  itggtgglgg  llarhlvtgh  gikhlllagr  rgpdapgara  lrdelaalga
       5051  evtvaacdva  draaldrlla  qlppehplta  vvhtagvldd  atvgtltper
       5101  ldtvlrakad  aawhlhdatr  drdlagfvly  ssvagvtggp  gqgnyaagnt
       5151  fldalaahra  aqglpglsla  wgpwgqdagm  tgtlgaadla  rlersgmppl
       5201  tpeqglalfd  aagargdgfa  vavrlargaa  apgadevpav  lralvrgrrr
  15   5251  taaaaghagv  larrlaalda  eqrhqalldl  vrtetaavlg  hsgadavpae
       5301  rdfnrlgfds  lmavelrtrl  atatgarlpa  tlvfdhptpd  avarhlastl
       5351  pggtaagpdr  splaeldria  aelspegadd  atrqgvvgrl  rhllaqwdgt
       5401  rqdgggttvd  drieaasaee  vlafidhelg  rqads
```

<u>SEQ ID NO. 39</u>

NysK

```
         1  mpdekklvdy  lkwvtkdlhq  trqrlqevea  grhepvaivg  macrfpggvr
        51  spedlwells  agrdgigpfp  adrgwdlaal  agdgpgrsat  qeggflpdaa
       101  afdpgffdis  prealamdpq  qrllletawe  aversgidpa  glrgsrtgvf
       151  vgtngqdyah  lvlaaqddmg  gyagnglaas  vlsgrlafal  glegpavtld
       201  tacssslvtl  hlaaqavrag  ecglalaggv  tvmttsssfa  gfslqgglap
       251  dgrckafaea  adgtgwsegi  glllverlsd  aqrnghpvla  vlrgsavnqd
       301  gasnglsapn  gpsqqrvirq  alagaglvpg  dvdaveahgt  gtrlgdpiea
       351  gallatygqd  rpadrplwlg  svksnlghtq  aaagvagvik  mvlalrhgvl
       401  pqtlhvdaps  shvdwesgav  rlltapvaws  egddrvrrag  vssfgisgtn
       451  ahvileqapd  qpeptaeeta  aaapggtaee  raaapvapra  vpwpvaarta
       501  galdaqlvrv  ralttapgrt  aadvghalat  artpfehral  lvheggavte
  35   551  vargavptgd  rgglavlfsg  qgsqrpgmgr  elharypvfa  aafdetvall
       601  darlgtslrd  ivwdqdrtrl  ddtrhtqpal  favevalyrl  laswgirpdh
       651  vtghsigeit  aahvagvltl  adactlvaar  atamselppg  gamvaleate
```

```
    701    devrplltdd  laiaavnapr  svvvagaeda  alavrrhfdd  lgrrttrlpv
    751    shafhsplmd  pmldafrtal  apltfaepei  pvvsnltglp  ataeelatph
    801    ywvchvrqav  rfgdgvrala  drgvrtflel  gpdgvlsalv  renlpepglv
    851    avpvlrkerp  eettvlaalg  tlwahgadvd  wdavfagtrt  pqadpvelpt
5   901    yafqrarywp  tlgarhgdpa  dlgqtaaahp  llgaavtlad  adetvltgrl
    951    alpshpwlgd  hrsdgritvp  gvafaelavr  agdlsgtphl  arldlpaplt
   1001    lgdgdtvtlq  vrvgapdpag  hrpltvharl  aatedapwtt  catgllapda
   1051    peapadpigp  adagwpprda  rpvpvadlda  aataagrhyg  phfqgltglw
   1101    rrdgevfaev  alptataadr  afgihpalla  talrataald  ddhtaghtpe
10  1151    ptgitglalh  atgatalrvr  ltatgpdtva  laaadatgga  vltadtvtlg
   1201    spqdrpapap  aghtgqgglf  hldwvpvdpg  sratgtrwav  vgddeldlgy
   1251    alhradetvs  ayaaslggai  gdsglapdvf  lvpvvggpda  gpdavhavta
   1301    ralgllqewl  neprlagarl  vfvtrgavav  pgetvtdpag  aavwgllrsa
   1351    qtenpgslll  vdlddafrsa  gmlphvltld  eqqlvvrdha  vraarlarlp
15  1401    epaagtapar  awdpdgtvli  tggtgglgaa  larhlvtvrg  arhlllagrr
   1451    gpeapgagel  vaeltaqgad  vrvaacdvgd  rtaldallat  vpaahpltav
   1501    vhtagvldda  ligsltpdql  atvlrpkada  awhlhdatrg  ldlagfvlys
   1551    svsgvlgspg  qgnyaaanay  ldalarhrad  qglpalslaw  gpwgrgsgmt
   1601    asvsdadler  margglpplt  vedglalfda  avgrpepalv  psrinvaglr
20  1651    dqqalpalwr  dlvprarrta  atadrspvtv  rerlrhldet  gqeqllidlv
   1701    vgytagllgh  pdptavdper  gflelgfdsl  vsvglrnqla  eilglrlpss
   1751    ivfdskspvk  larwlhqela  ngpqpgatgp  aaadarpavr  sddtleglfy
   1801    navrggklve  amrmlkavan  trpmfdtpae  leelsepvtl  adgpgrprli
   1851    fvsapgatgg  vhqyariaah  frgsrhvsal  plmgfapgel  lpatseaaar
25  1901    ivaesvlmas  egepfvmvgh  stggslayla  agvledtwdv  rpeavvlldt
   1951    asirynpgeg  ndldrttrfy  ladidspsvt  lnsarmsama  hwfmamtdiq
   2001    apaptaptll  vraaraldgf  rldtssvpad  evrdidadhl  slakehsalt
   2051    aqaiegwlae  lpdpaa
```

30   SEQ ID NO. 40

NysL


```
     1    mstptappsl  kaevppvlrl  spllrelqsr  apvckvrtpa  gdegwlvtrh
35   51    telkqllhdd  rlarahadpa  napryvhnpf  ldllvvddfd  lartlhaemr
    101    slftpqfsar  rvmdltprvc  alaegvlahf  vaqgppadlh  ndfslpfsls
    151    vlcaligvpa  eeqgkliaal  tklgelddpa  rvqegqdelf  gllsglarrk
```

```
201  ritpeddvis  rlclkvpsde  rigpiasgll  fagldsvash  idlgtvlfiq
251  hpdqlaaala  deklmrgave  eilrsakagg  svlpryatad  vpigdvtira
301  gdlvlldftl  vnfdrtvfde  pelfdirrap  nphltfghgm  whcigaplar
351  vnlrtaytll  ftrlpglrlv  rpveelrvls  gqlsagltel  pvtw
```

SEO ID NO. 41

NysM


```
  1  vritvdpgrc  vgagqcvlta  pdlfdqdddg  lvtvlagaad  aadpgdvrda
 51  aalcpsgais  vaad
```


SEO ID NO. 42

NysN


```
  1  msteadarta  apqcpvafpl  rrpgrpfppp  eyatyrggag  lvrselpsgp
 51  vwlvtrhedv  ravltdpris  adpskpgfpk  agrtggapsq  yevpgwfvam
101  dppehgrfrk  tlipeftvrk  vrelrpviqq  ivderidaml  aagtsadlve
151  sfalpvpslv  issllgvpkv  drdffedrtr  vlvrlsstde  erdkatqall
201  rylgrliqik  qrrpgddlis  rliaagtlsr  qelsgvamll  liaghettan
251  niglgvvqll  tnprwigddr  iveellryys  vadlvafrva  vedveiggql
301  iragegivpl  iaaanhdata  faapsefdpe  rsarshvafg  ygvhqclgqn
351  lvreemdiay  rtlfaripsl  tlavpveelp  lkydgvlfgl  helpvtwk
```


SEO ID No. 43:

Nys RIV (long) 266 aa

VTITHLTDSNHRTTGGVISAQTAPAGESVGPGLMASLDRDLTIKHANQEFRRRFD
DSAGDVCGRSFRDLMHPSVQQPLMRQFSRLIEGKRHRFASHVVAVGAQDAAFAGT
LTASAVTGKTPDIAGILVLMDSSGAADAADAGVVTSQKKFLTEIDARILEGIAAG
LSTIPLASRLYLSRQGVEYHVTGLLRKLRVPNRAALVSRAYSMGILNVGTWPPKV
VDDFIK

SEO ID NO. 44:
GCC GGC ATG CGA CGA ACA GGA CGA GAG GT

SEQ ID NO. 45:
GCC GTG GTC GAC GAA GGC

SEQ ID NO. 46:
5   CTC AGC ATG CCG AAA GGA TGG CGG

SEQ ID NO. 47:
AGG CAA GCT TCG GCG ACA CGG GCG T

10   SEQ ID NO. 48:
CTC AGC ATG CGT ACG ACC GGC GGG

(19) World Intellectual Property Organization
International Bureau

(51) International Patent Classification[7]: C12N 15/52, C12P 17/18, 19/62, C07H 17/08

(21) International Application Number: PCT/GB01/00509

(22) International Filing Date: 8 February 2001 (08.02.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0002840.7   8 February 2000 (08.02.2000)   GB
0008786.6   10 April 2000 (10.04.2000)   GB
0009387.2   14 April 2000 (14.04.2000)   GB

(71) Applicants (for all designated States except US): NORGES TEKNISK NATURVITENSKAPELIGE UNIVERSITET [NO/NO]; N-7491 Trondheim (NO). STIFTELSEN FOR INDUSTRIELL OG TEKNISK FORSKNING VED NORGES TEKNISKE HØGSKOLE [NO/NO]; Sintef Applied Chemistry Division, Sem Seelands Vel 2A, N-7034 Trondheim (NO).

ALPHARMA AS [NO/NO]; Harbitzalléen 3, N-0212 Oslo (NO). SINVENT AS [NO/NO]; Strindveien 4, N-7465 Trondheim (NO).

(71) Applicant (for GB only): DZIEGLEWSKA, Hanna [GB/GB]; Frank B. Dehn & Co., 179 Queen Victoria Street, London EC4V 4EL (GB).

(71) Applicants and
(72) Inventors: ZOTCHEV, Sergey Borisovich [RU/NO]; Elgester gate 51, N-7030 Trondheim (NO). SEKUROVA, Olga Nikolayivna [RU/NO]; Elgester gate 51, N-7030 Trondheim (NO). FJÆRVIK, Espen [NO/NO]; Lianveien 10, N-7025 Trondheim (NO). BRAUTASET, Trygve [NO/NO]; Martin Stokkens vei 74, N-7027 Trondheim (NO). STRØM, Arne, Reidar [NO/NO]; Konvallvegen 38, N-7050 Trondheim (NO). VALLA, Svein [NO/NO]; Naustanveien 2, N-7560 Vikhamar (NO).

(72) Inventors; and
(75) Inventors/Applicants (for US only): ELLINGSEN, Trond, Erling [NO/NO]; Markaplassen 273, N-7054 Ranheim (NO). SLETTA, Håvard [NO/NO]; Uglaveien

(54) Title: GENE CLUSTER ENCODING A NYSTATIN POLYKETIDE SYNTHASE AND ITS MANIPULATION AND UTILITY
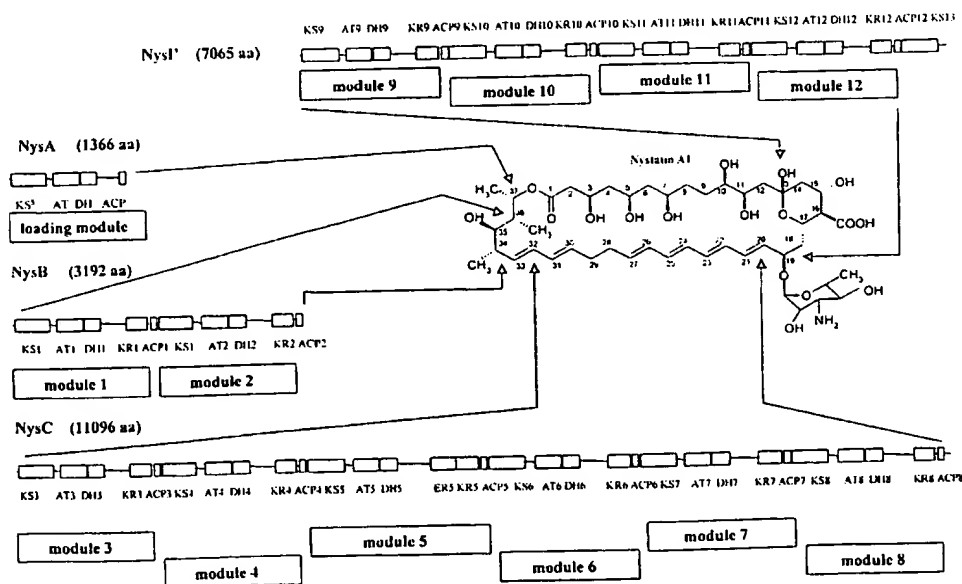
(57) Abstract: The invention provides a nucleic acid molecule comprising: a nucleotide sequence as shown in SEQ ID No. 35; wherein said sequence preferably encodes or is complementary to a sequence encoding a nystatin PKS enzyme or a part thereof. Also provided are part of such molecules and polypeptides (and parts thereof) encoded by such a nucleic acid molecule, and the use of such molecules and polypeptides in facilitating nystatin biosynthesis and in the synthesis of nystatin derivatives and novel polyketide as macrolide structures.

19, N-7024 Trondheim (NO), GULLIKSEN, Ole-Martin [NO/NO]; Grevlingvn. 31 D, N-0595 Oslo (NO).

(74) **Agents: DZIEGLEWSKA, Hanna** et al.; Frank B. Dehn & Co., 179 Queen Victoria Street, London EC4V 4EL (GB).

(81) **Designated States** *(national)*: AE, AG, AL, AM, AT, AT (utility model), AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, CZ (utility model), DE, DE (utility model), DK, DK (utility model), DM, DZ, EE, EE (utility model), ES, FI, FI (utility model), GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SK (utility model), SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) **Designated States** *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
    with international search report

(88) **Date of publication of the international search report:**
14 March 2002

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

Ir ational Application No

PCT/GB 01/00509

A. CLASSIFICATION OF SUBJECT MATTER

IPC 7   C12N15/52   C12P17/18   C12P19/62   C07H17/08

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7   C12N   C12P   C07H

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, EMBL, WPI Data, PAJ, BIOSIS, MEDLINE, EMBASE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE EM_PRO 'Online! EMBL; ID SNA132222, AC AJ132222, 20 April 1999 (1999-04-20) APARICIO J F: "Streptomyces natalensis pimS1 gene" XP002174375 Note: 60.5%, 60.4% and 60.5% nt seq identity with SEQ ID NO:35, 1 and 2 in 6853, 5454 and 6853 nt overlap (5017-11675:6892-13554, 5974-11277:29502-34804, and 5017-11675:6892-13554), respectively page 4-5	1-7,9, 14,17, 19-24, 27,30-36
X	-& APARICIO J F ET AL.: "The biosynthetic gene cluster for the 26-membered ring polyene macrolide pimaricin: A new polyketide synthase organization encoded by two subclusters separated by _/--	1-7,9, 14,17, 19-24, 27,30-36

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

'A' document defining the general state of the art which is not considered to be of particular relevance

'E' earlier document but published on or after the international filing date

'L' document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

'O' document referring to an oral disclosure, use, exhibition or other means

'P' document published prior to the international filing date but later than the priority date claimed

'T' later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

'X' document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

'Y' document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

'&' document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
8 August 2001	22/08/2001

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	van de Kamp, M

3

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
	functionalization genes." JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 274, no. 15, 9 April 1999 (1999-04-09), pages 10133-10139, XP002174372 abstract page 10139, left-hand column, line 12-14 ---	
X	DATABASE EM_PRO 'Online! EMBL; ID SCI7, AC AL096743, 1 July 1999 (1999-07-01) SEEGER K ET AL.: "Streptomyces coelicolor cosmid I7" XP002174376 Note: 68.3% nt seq identity (71.3% ungapped) with SEQ ID NO:35 and 1 in 1077 nt overlap (13366-14424:121753-122803 and 13366-14424:61492-62542, respectively) page 6, line 70 -page 7, line 18 page 16 ---	1-6,14, 15
X	DATABASE SWALL 'Online! SWISSPROT; ID Q9X9X5, AC Q9X9X5, 1 November 1999 (1999-11-01) SEEGER K ET AL.: "PUTATIVE TRANSCRIPTIONAL REGULATOR" XP002174377 Note: 65.6% aa seq identity with SEQ ID NO:13 in 250 aa overlap (9-258:4-253) the whole document ---	1-6,14, 15
X	SHENIN Y D ET AL.: "Nystatin: Methods for preparation, search for derivatives, and prospects for application in medicine (review)." KHIMIKO-FARMATSEVTICHESKII ZHURNAL, vol. 27, no. 2, 1993, pages 14-21, XP000984316 ISSN: 0023-1134 the whole document figure on page 17 ---	35,36
X	US 5 981 721 A (MOHAN ARTHUR G) 9 November 1999 (1999-11-09) the whole document example 12 ---	35,36
	-/--	

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
X	MCDANIEL R ET AL: "Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel unnatural natural products" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, vol. 96, 1999, pages 1846-1851, XP000910246 ISSN: 0027-8424 abstract; figure 2 ---	32,33
Y	MALPARTIDA F ET AL: "Homology between Streptomyces genes coding for synthesis of different poyletides used to clone antibiotic biosynthetic genes" NATURE (LONDON), vol. 325, no. 6107, 1987, pages 818-821, XP002075972 ISSN: 0028-0836 the whole document ---	1-28, 30-36
Y	DECKER H ET AL.: "A general approach for cloning and characterizing dNDP-glucose dehydratase genes from actinomycetes." FEMS MICROBIOLOGY LETTERS, vol. 141, no. 2-3, 1996, pages 195-201, XP001010562 ISSN: 0378-1097 the whole document ---	1-28, 30-36
Y	KHOSLA C ET AL: "Generation of polyketide libraries via combinatorial biosynthesis" TRENDS IN BIOTECHNOLOGY, vol. 14, no. 9, 1 September 1996 (1996-09-01), pages 335-341, XP004035738 ISSN: 0167-7799 the whole document ---	23-28, 30-36
A	LEONARD KATZ: "Manipulation of modular polyketide synthases" CHEMICAL REVIEWS, vol. 97, no. 7, 1997, pages 2557-2575, XP002103748 ISSN: 0009-2665 paragraphs I and VII ---	

-/--

3

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
A	SEKUROVA O ET AL: "Molecular cloning and analysis of a pleiotropic regulatory gene locus from the nystatin producer Streptomyces noursei ATCC11455." FEMS MICROBIOLOGY LETTERS, vol. 177, no. 2, 15 August 1999 (1999-08-15), pages 297-304, XP001015583 ISSN: 0378-1097 cited in the application abstract	
A	VANDEN BOSSCHE H ET AL: "Anti-Candida drugs - the biochemical basis for their activity" CRITICAL REVIEWS IN MICROBIOLOGY, vol. 15, no. 1, 1987, pages 57-72, XP001015427 ISSN: 0045-6454 paragraphs I and II	
P,X	BRAUTASET T ET AL.: "Biosynthesis of the polyene antifungal antibiotic nystatin in Streptomyces noursei ATCC 11455: Analysis of the gene cluster and deduction of the biosynthetic pathway." CHEMISTRY & BIOLOGY, vol. 7, no. 6, June 2000 (2000-06), pages 395-403, XP000953274 ISSN: 1074-5521 the whole document	1-28, 34-36
P,X	WO 00 77222 A (COLINA ANGEL JOSE ;MARTIN JUAN FRANSISCO (ES); APARICIO JESUS F (E) 21 December 2000 (2000-12-21) the whole document page 18, line 5-17 examples 1-4 figures 3,4	1-7,9, 14,17, 19-24, 27,34-36
T	ZOTCHEV S ET AL.: "Identification of a gene cluster for antibacterial polyketide-derived antibiotic biosynthesis in the nystatin producer Streptomyces noursei ATCC 11455." MICROBIOLOGY, vol. 146, no. 3, March 2000 (2000-03), pages 611-619, XP001012427 ISSN: 1350-0872 the whole document	

3

INTERNATIONAL SEARCH REPORT

Information on patent family members

Patent document cited in search report		Publication date	Patent family member(s)		Publication date
US 5981721	A	09-11-1999	NONE		
WO 0077222	A	21-12-2000	AU	5407400 A	02-01-2001

Form PCT/ISA/210 (patent family annex) (July 1992)